Production of mammalian glycoproteins for structural analysis: site-specific recombination systems in CHO cells

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina

zu Braunschweig

zur Erlangung des Grades einer

Doktorin der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Sonja Wilke
aus Peine

1. Referentin oder Referent:	Honorarprofessor Dr. Dirk Heinz
2. Referentin oder Referent:	Professor Dr. Stefan Dübel
eingereicht am:	03.03.2011
mündliche Prüfung (Disputation) am:	07.07.2011

Druckjahr 2011

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikation

Wilke, S., Krausze, J., Gossen, M., Groebe, L., Jäger, V., Gherardi, E., van den Heuvel, J. & Büssow, K. Glycoprotein production for structure analysis with stable, glycosylation mutant CHO cell lines established by fluorescence-activated cell sorting. Protein Science 19, 1264 – 1271 (2010).

Wilke, S., Groebe, L., Jäger, V., Josewski, J., Duda, A., Polle, L., Gossen, M., Wirth, D., van den Heuvel, J., Heinz, D. & Büssow K. Streamlining Glycoprotein Crystallization by Targeted Development of Production Cell Lines (submitted).

Tagungsbeiträge

Wilke, S., Krausze, J., Gossen, M., Groebe, L., Jäger, V., Gherardi, E., van den Heuvel, J. & Büssow, K.: Producing mammalian proteins for structures: site-directed recombination in CHO Lec3.2.8.1 cells. (Vortrag) ACTIP Meeting, Penzberg (2010). Fellowship Award Winner.

Wilke, S., Tokarski, S., Jäger, V., van den Heuvel, J., Gossen, M., Gherardi, E. & Büssow, K.: New ways of establishing production cell lines for structural biology. (Poster) 21st ESACT Meeting, Dublin, Ireland (2009). First Poster Prize.

Wilke, S., Tokarski, S., Jäger, V., van den Heuvel, J., Gossen, M., Gherardi, E. & Büssow, K.: Establishing mammalian production cell lines for structural biology by site-specific recombination. (Vortrag, Poster) 9th PEACe Conference, Jackson, Wyoming, USA (2009).

Wilke, S., Tokarski, S., Jäger, V., Reichelt, J., Klink, B. & Büssow, K.: X-Ray analysis of the human lysosome-associated membrane protein 3. (Poster) M2M-9: From Measurement to Model, EMBL Hamburg, Germany (2009).

Für Max und Maria Teichmann

Content

Abbrevations

aa	Amino acids
Å	Ångstrøm (1 Å = 0.1 nm)
A_λ	Absorption at the wavelength λ in nm (equivalent to OD_λ)
Ab	Antibody
APS	Ammonium persulphate
ATCC	American Type Culture Collection
ATG	Translation start codon
ATP	Adenosine triphosphate
BESSY	Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung
BLAST	Basic local alignment search tool
bp	Base pair
BSA	Bovine serum albumin
c	Concentration
CHO	Chinese hamster ovary
CMV	Cytomegalovirus
Cre	Cyclization recombination
Da	Dalton
DC-LAMP	Dendritic cell Lysosome-associated membrane protein
DESY	Deutsches Elektronen-Synchrotron
DLS	Dynamic light scattering
DMEM	Dulbecco's modified Eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleosid-triphosphate
DTT	Dithiothreitol
E	PCR efficiencies
E. coli	*Escherichia coli*
ECL	Enhanced chemiluminescence
EDTA	Ethylen diamine tetra acetic acid
EF1-α	Elongation factor 1 α
EGFP	Enhanced green fluorescent protein
ELISA	Enzyme-linked immunosorbent assay

EMBL	The European Molecular Biology Laboratory
ESI	Electrospray ionisation
EtOH	Ethanol
FACS	Fluorescence-activated cell sorter
FCS	Fetal calf serum
FLEx	Flp-mediated excision
Flp	Flippase
FRT	Flp recognition target
FSC	Forward scatter
G418	aminoglycoside-2'-deoxystreptine (Gentamycin-derivative)
GalNAc	N-acetylgalactosamine
GFP	Green fluorescent protein
GlcNAc	N-acetyl-glucosamine
GPC	Gel permeation chromatography
h	Hour
HEK	Human embryonic kidney
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HGF	Hepatocyte growth factor
His	Histidine
His_6	6 consecutive histidine residues
HPSF	High Purity Salt Free
HRP	Horseradish peroxidise
Hyg	Hygromycin-B-Phosphotransferase
HygTk	Fusion of Hyg and Tk
IRES	Internal ribosome entry site
L	Liter
LAMP	Lysosome-associated membrane protein
hLAMP3-prox	Human luminal membrane-proximal domain of LAMP-3
LB	Luria Bertoni
LoxP	loxP locus of X-over (target sequence for the Cre recombinase)
MAD	Multiple anomalous dispersion
mAb	Monoclonal antibody
MALDI	Matrix assisted laser desorption / ionisation
MCS	Multiple cloning site

MeOH	Methanol
MS	Mass spectrometry
MW	Molecular weight
N-	Amino terminal
Neo	Neomycin phosphotransferase
No.	Number
o.n.	Over night
ori	Origin of replication
PAGE	Polyacrylamide gelelectrophoresis
PBS	Phosphate buffered saline
pcd	Picogram per cell per day
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PEI	Polyethylenimine
PGK	Phosphoglycerate kinase
PI	Propidium iodide
polyA	Polyadenylation signal
PVDF	Polyvinylidene difluoride
RMCE	Recombinase-mediated cassette exchange
r.m.s.d.	Root mean square deviation
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Rotations per minute
RT	Room temperature
scHGF	Single chain hepatocyte growth factor
SDS	Sodium dodecyl sulphate
SeMet	Selenomethionine
SSC	Side scatter
Tm	Melting temperature
TBS	Tris buffered saline
TBST	TBS containing 0.1 % (v/v) TWEEN-20
TEMED	N, N, N', N'-Tetramethylethylenediamine
TMB	3,3',5,5'-tetramethylbenzidine

Tk	Thymidine kinase
TOF	Time of flight
Tris	Tris-(hydroxymethyl)-aminomethane
v/v	Volume-volume percentage
w/v	Weight-volume percentage
w/w	Weight-weight percentage
wt	Wild type

Zusammenfassung

Das Interesse an Raumstrukturen glykosylierter Proteine hat in den letzten Jahren stetig zugenommen, da diese Moleküle Schlüsselfunktionen im menschlichen Körper wahrnehmen und unter ihnen viele wichtige pharmakologische Zielstrukturen sind. Aufgrund ihrer Fähigkeit zur posttranslationalen Modifikation und komplexen Faltung stellen stabile Zellinien ein hervorragendes System für die Produktion rekombinanter Glykoproteine dar. Obwohl die Proteinproduktion in Säugerzellen schon wesentlich weiterentwickelt wurde, bleibt die Etablierung stabiler Zelllinien kompliziert und zeitaufwendig. Die zufällige Transgenintegration ins Wirtsgenom erfordert ein aufwendiges Überprüfen sehr vieler Zellklone. Ein weiteres Problem, insbesondere für die Röntgenstrukturanalyse, ist die Heterogenität und Flexibilität der Protein-Glykosylierungen, die häufig die Bildung von gut streuenden Kristallen verhindern. Mutierte Zelllinien wie CHO Lec3.2.8.1 produzieren Glykoproteine mit verkürzten Kohlenhydrat-Seitenketten vom mannose-reichen Typ (*high mannose)*. Diese Glykoproteine können deglykosyliert und effizient kristallisiert werden. Um zu gut kristallisierbaren Proteinkonstrukten zu gelangen, müssen oft unterschiedliche Varianten von einem Protein hergestellt werden. Daher ist ein schnelles und flexibles Expressionssystem gerade für die Strukturbiologie von großem Interesse.

Diese Arbeit beschäftigt sich mit gezielter Genmanipulation mit Flp Rekombinase-vermittelten Rekombinationsstrategien, um die Zelllinienentwicklung in CHO Lec3.2.8.1 Zellen im Hinblick auf einen höheren Durchsatz und vorhersagbare Expressionsstärke zu optimieren. Zuerst wurden genetisch stabile Loci mit starker Genaktivität im Wirtszellgenom mit GFP Kassetten markiert, die von Flp-Erkennungssequenzen (FRT) flankiert waren. Zwei aufeinander folgende Fluoreszenz-aktivierte Zellsortierungsschritte (FACS) ermöglichten die Isolierung und Klonierung der am stärksten fluoreszierenden Zellen. Zellen, die GFP über mehrere Wochen stabil exprimierten, wurden Masterzellklone genannt. Je nach Strategie wurde das GFP entweder über Rekombinase-vermittelten Kassettenaustausch (RMCE) gegen ein anderes Gen ausgetauscht oder es wurde über ortsgerichtete Rekombination ausgeschnitten (FLEx). Im letzteren Fall war das zu exprimierende Gen schon hinter das GFP-Gen kloniert worden. Durch Entfernen des GFPs wurde das zu exprimierende Gen unter die Kontrolle des Promoters gebracht und damit aktiviert. Ein Vergleich der beiden Rekombinationssysteme zeigte die höhere Flexibilität und Schnelligkeit des RMCE Systems. Etablierte RMCE Masterzellen können in einem Schritt innerhalb von 2

Monaten umgewandelt werden, um beliebige Proteine zu produzieren. Auf diese Weise wurden 10 unterschiedliche Produktionszelllinien aus den Masterzellen generiert. Die rekombinanten Produkte umfassten hauptsächlich sekretierte Glykoproteine, aber auch ein intrazelluläres Protein und ein GPI-Anker-Protein. Drei der Zelllinien wurden in 2,5 L Bioreaktoren kultiviert, wobei das jeweilige rekombinante Protein reproduzierbar im Milligramm-pro-Liter Maßstab produziert wurde.

Eines der Zielproteine, eine Domäne des humanen LAMP-3/DC-LAMP, ein Mitglied der Familie der lysosomal-assoziierten Membranproteine, wurde gereinigt, deglykosyliert und für die Röntgenstrukturanalyse kristallisiert. Schwermetallionen-Derivate der Kristalle wurden verwendet, um das Phasenproblem über anomale Dispersion mit multiplen Wellenlängen (MAD) zu lösen. Diese Experimente ermöglichten es, die Proteinstruktur aufzuklären.

Zusammengefasst ermöglicht die Kombination aus FACS und ortsspezifischer Rekombination eine schnelle und robuste Klonierung von Zelllinien, die das rekombinante Protein ohne ständige Antibiotika-Selektion stabil produzieren. Aufgrund der kurzen Etablierungszeit ist das RMCE System in CHO Lec3.2.8.1 Zellen das ideale System, um Glykoproteine für die Strukturaufklärung zu produzieren.

Summary

Structures of mammalian glycoproteins attracted increasing interest in recent years, as these proteins have key functions in the human body and include important drug targets. Due to their capability of post-translational modification and complex folding, stable cell lines are an excellent expression system for their recombinant production. Though considerable enhancements improved the production in mammalian cell lines, their development remains difficult and time-consuming. Transfections leading to random transgene integration require exhaustive clone screening. Another challenge, in particular for X-ray structure analysis, is the heterogeneous and flexible nature of protein glycosylation which often hampers the formation of well diffracting crystals. Glycosylation mutant cell lines such as CHO Lec3.2.8.1 produce glycoproteins with truncated carbohydrate chains of high-mannose type. These glycoproteins can be deglycosylated and crystallized efficiently. To find well crystallisable protein constructs, structural projects often necessitate the production of a variety of different constructs per protein. Thus, a fast and flexible expression system in CHO Lec3.2.8.1 cells is of essential interest for structural biology.

This work describes defined genome engineering by Flp-mediated recombination strategies in order to optimise cell line development in CHO Lec3.2.8.1 towards higher-throughput and predictable expression strength. Firstly, stable and high-producing loci in the host genome were tagged by FRT-flanked GFP cassettes. Two subsequent rounds of preparative fluorescent activated cell sorting (FACS) enabled high-throughput selection by isolating and cloning the top fluorescent cells. Cells stably expressing GFP over several months were termed master cell clones. Depending on the strategy, the GFP gene of these cells was either exchanged against another gene by recombinase-mediated cassette exchange (RMCE) or excised by site-specific recombination (FLEx). In the latter case, the gene of interest was already located downstream of GFP. Upon deletion by recombination, the gene of interest was put under control of the promoter and thus expressed. A comparison of both recombination systems proved a higher flexibility and speed of RMCE. Once established, RMCE master cell lines can be converted to produce any protein of interest in about two months. Thus, 10 different production cell lines were successfully established from selected master cells. The recombinant products include glycoproteins, an intracellular protein and a GPI-anchored protein. Three cell lines were cultivated in 2.5 L bench top bioreactors and reliably produced recombinant protein in mg per litre scale.

As an example, one of these proteins, a domain of human LAMP-3/DC-LAMP, a member of the lysosome associated membrane protein family, was purified, deglycosylated and crystallized for X-ray structure analysis. Highly diffracting heavy-atom derivative crystals were used to obtain phase information by multi-wavelengths anomalous dispersion (MAD) experiments. This information enabled the determination of the protein structure.

In conclusion, the combination of FACS and site-specific recombination facilitates rapid and robust cloning of cell lines that stably produce recombinant proteins without continuously adding antibiotics. Due to its short development timeline, RMCE in CHO Lec3.2.8.1 cells is the ideal system to produce glycoproteins for structure determination.

1 Introduction

1.1 Mammalian expression systems in structural biology

Three-dimensional structures at atomic resolution are indispensable for a thorough understanding of protein function. In particular, the identification of functional sites and binding partners of proteins involved in diseases facilitates pharmaceutical drug development (Buchanan, 2002; Zhang & Kim, 2003). The first bottleneck for structure determination is the availability of sufficient amounts of soluble, high-quality protein. According to structures deposited in the Protein Data Bank (PDB), bacteria remain the most common expression hosts, assigned to 88% of the PDB entries, followed by insects/baculovirus systems (4%), fungi/yeasts (2.6%) and mammalians (2.4%) (Nettleship *et al.*, 2010). Although constituting a valuable high-throughput system, bacteria lack critical eukaryotic post-translational modifications like glycosylation and certain chaperones to ensure a proper folding of many mammalian proteins including glycoproteins, membrane proteins and large multi-protein complexes. The demand for eukaryotic expression systems capable of dealing with the production of these challenging proteins is further underlined by the estimation that about 50% of all human proteins are glycosylated (Apweiler *et al.*, 1999). Due to authentic biosynthesis and folding, mammalian expression systems are favoured for proteins of the same origin (Wurm, 2004). However, in comparison to other eukaryotic expression systems mammalian cells are generally regarded as a time- and cost-intensive choice. Their cultivation is characterized by slow growth kinetics, low cell densities, high sensitivity to mechanical stress and the need for specialized media (Sandig *et al.*, 2005). Progress in high-throughput techniques and in genetic engineering has made mammalian cell culture more attractive in terms of speed, ease and yield (Andersen & Krummen, 2002). Hence, the number of protein structures based on mammalian expression systems has progressively increased each year and is poised to get ahead of fungi/yeast expression systems (Fig. 1-1). Insect/ baculovirus expression systems remain the most abundant in eukaryotic systems. Comparative studies on the performance of baculovirus- and mammalian cell based systems stated that their expression abilities complement one another (Aricescu *et al.*, 2006). While cytosolic proteins were favourably produced in baculovirus expression systems, secreted proteins were obtained in higher yields in mammalian cells.

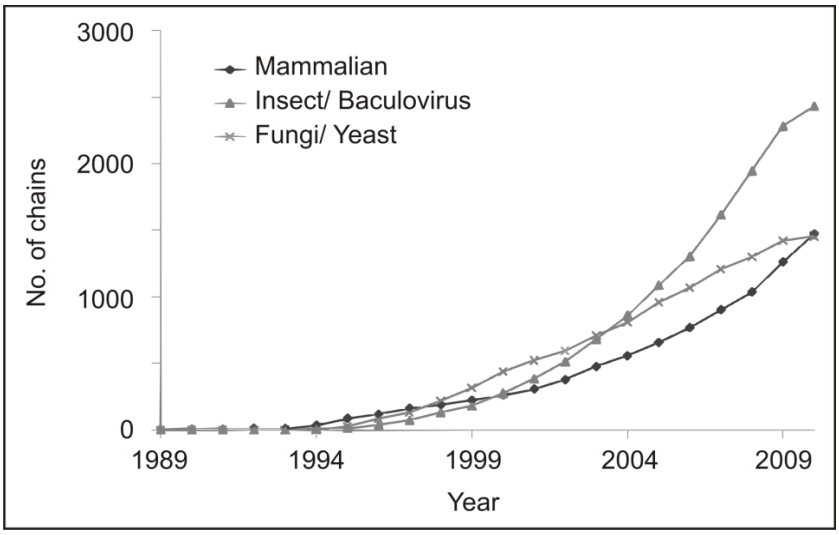


Figure 1-1: Plot of cumulative total number of protein chains deposited in the PDB whose expression system was identified as mammalian, insect, baculovirus, fungi or yeast by year of deposition. Expression data were parsed from the set of PDB files available in August 2010. Chains were counted rather than PDB entries as expression information is recorded by chains in the PDB.

The most common mammalian expression hosts for structural studies are human embryonic kidney (HEK293) cells (Graham *et al.*, 1977) and Chinese hamster ovary (CHO) cells (Tjio & Puck, 1958) (Fig. 1-2). Transient expression is distinguished from stable cell lines. Transient expression is characterized by a straightforward process from DNA delivery to protein harvest within a few days. By the introduction of suspension cultures and improvements in large-scale transfection techniques, transient expression systems have gained new advantages regarding speed and throughput (Baldi *et al.*, 2007; Derouazi *et al.*, 2004; Durocher *et al.*, 2002; Girard *et al.*, 2002; Meissner *et al.*, 2001; Pham *et al.*, 2003). Due to their high transfection rates, HEK293 cells are valuable hosts for transient expression and have successfully delivered protein for various X-ray structure determinations (Bishop *et al.*, 2009; Carafoli *et al.*, 2008). To enhance protein expression further, two genetic variants, HEK293T and HEK293E, were developed. While HEK293T cells constitutively express the simian virus 40 (SV40) large T-antigen (Lebkowski *et al.*, 1985), HEK293E cells express the Epstein-Barr virus (EBV) nuclear antigen 1 (EBNA1) (Cachianes *et al.*, 1993). Both gene products enable the episomal amplification of plasmids containing the corresponding viral SV40 (HEK293T) or EBV (HEK293E) origin of replication (Van Craenenbroeck *et al.*, 2000). Recently, a comparable system for episomal replication and plasmid retention was developed in CHO cells (Kunaparaju *et al.*, 2005). This so-called Epi-CHO expression system consists of CHO-K1 cells stably producing polyomavirus (Py) large T-antigen (PyLT) and a transfection vector bearing the Py origin of replication for autonomous

amplification. The transfection vector additionally encodes the EBNA1 gene and the EBV origin of replication for plasmid retention. Despite these enhancements, large DNA quantities have to be prepared regularly for each transfection. Thus, low-yield recombinant proteins are better expressed by stable cell lines which can be scaled up more easily. Moreover, their cultivation can be repeated indefinitely without the need of large DNA quantities. In pharmaceutical industry, stable mammalian cell lines are the dominant system for the production of recombinant therapeutic proteins by supplying 60 to 70% of these products (Wurm, 2004). CHO cells are the most popular hosts for efficient integration of exogenous DNA in their genome (Hoeijmakers *et al.*, 1987).

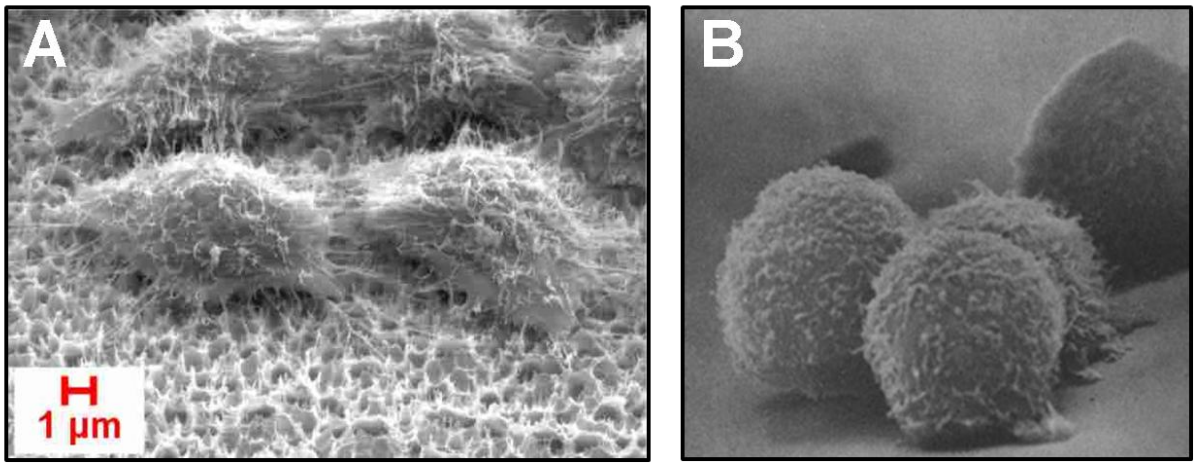


Figure 1-2: Electron scanning microscopy of HEK293 and CHO cells. (A) HEK293 cells on silicon surface (http://www.fz-juelich.de). (B) Non-adherent CHO cells (Porter *et al.*, 1973).

1.2 Conventional cell line development

1.2.1 The general strategy for cell line development

The development of a production cell line is conventionally based on random integration of a transgene into the host cell genome (Fig. 1-3). In addition to the transgene, a selectable marker gene is co-transferred into the cells, either encoded on the same vector or on a separate vector (Gurtu *et al.*, 1996). The selection procedure depends on the marker which includes, for instance, the bacterial neomycin phosphotransferase, conferring resistance to the drug G418 (Southern & Berg, 1982), dihydrofolate reductase (DHFR) (Urlaub & Chasin, 1980) and glutamine synthetase (GS) (Cockett *et al.*, 1990). The DHFR and GS selection systems are based on the complementation of a deficient nucleotide metabolism and the glutamine synthesis pathway, respectively. Host cells are available for both systems that have a defect in the respective enzyme.

However, such a host is not necessarily needed for the GS system. Non-transformed cells are prevented from growth in the absence of the appropriate metabolites hypoxantine and thymidine (DHFR system) or glutamine (GS system). These metabolic selection systems are also suitable for gene amplification leading to increased transgene expression. To this end, the cells are cultured with an inhibitor of DHFR (methotrexate, MTX) or GS (methionine sulfoximine, MSX). Consequently, only cells with enhanced enzyme expression survive due to a transcriptionally active integration locus or increased vector copy numbers. A stepwise increase of the inhibitor can result in several hundred to thousand copies (Wurm *et al.*, 1986). The enhanced DHFR or GS production is usually accompanied by enriched product yields (Kim *et al.*, 2001). Since the specific productivity varies among individual clones, repeated selection and analysis cycles are necessary, making this procedure laborious and not suitable for all cell lines.
Following selection and eventual gene amplification, single cells are isolated, subcloned and screened for homogenous high-producers. Suitable cell clones are expanded, characterised and used for the process development destined to optimise cultivation conditions for higher productivity. Altogether, developing a cell line from gene transfer to larger-scale production can take more than a year and has to be repeated for each novel recombinant product (Wurm, 2004). In particular, the screening for stable high-producers is laborious, as expression abilities can vary essentially among transformed subclones (Gellissen, 2005). Only a minor part expresses the transgene in high concentration over a long time. Thus, many clones have to be analysed over a long time to find an ideal producer.

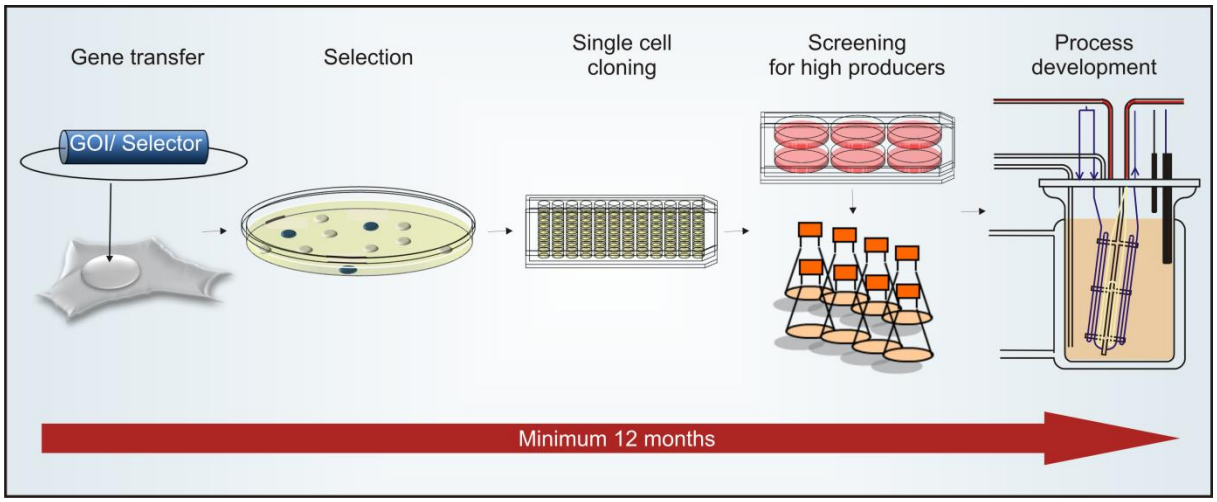


Figure 1-3: Conventional strategy for stable cell line development. Initially, a transgene and a selection marker are introduced to the host cell line by gene transfer methods. Cells having integrated the exogenous DNA in their genome are selected and cloned as individual cell lines. After freezing cell banks in liquid nitrogen, cell lines are screened for stable and high producers. For stable high-producers a process development confers efficient cultivation.

Factors influencing the transgene expression on genomic level are the number of integrated transgene copies and the genetic environment of the integration site which is known as position effect variegation (Bestor, 2000; Kito *et al.*, 2002). While integration into a heterochromatic locus leads to transcriptional down regulation culminating in complete transgene silencing, loci that can mediate transcriptional stabilization are rare (Pikaart *et al.*, 1998; Walters *et al.*, 1996). In addition, removing antibiotic selection pressure during cell propagation frequently results in mosaic gene silencing and diminishing protein yields (Liu *et al.*, 2006). Although an increased vector copy number upon gene amplification coincided with higher expression levels, gene copies at different loci tended to recombine and thus caused chromosomal aberration (Derouazi *et al.*, 2006). Tandem-repeat integrations are similarly problematic, as they may induce epigenetic inactivation as a consequence of a natural defence against retroviruses (McBurney *et al.*, 2002). Another defence strategy to protect the mammalian genome from foreign DNA sequences was described by Scrable and Stambrook (Scrable & Stambrook, 1997). While a *lacI* gene from *E. coli* was inactivated by methylation in transgenic mice, a synthetic version of the same gene that followed human codon usage was relatively resistant to methylation and silencing. Thus, prokaryotic features of the transfection vector which are required for plasmid replication and selection during cloning procedures may also influence transcriptional activity of the integration locus due to their divergent nucleotide composition.

1.2.2 Epigenetic strategies to augment transgene expression

The growing understanding of the molecular network that drives transcription of recombinant proteins gave rise to a variety of approaches employing epigenetic knowledge to increase product yield of stable cell lines. With regard to the position effect, one of the oldest strategies to create and maintain a favourable chromatin conformation is based upon the differences in histone modifications between the active and repressed status. While euchromatin is generally associated with histone acetylation, repression involves the removal of acetyl groups by histone deacetylases (HDACs). Hence, blocking HDACs by adding sodium butyrate to the cultured cells enhances transcriptional activity (Palermo *et al.*, 1991). A disadvantage of sodium butyrate is the inhibition of cell growth and induction of cellular apoptosis (Davie, 2003).

Further strategies involve genetic engineering of the host cells. The incorporation of *cis*-acting DNA sequences into the transgene vector was found to create transcriptional hot spots independent of the site of integration. Common examples of these regulatory elements are scaffold/matrix attachment regions (S/MARs) (Girod *et al.*, 2005; Kim *et al.*, 2004), which affect chromatin structure, and ubiquitous chromatin opening elements (UCOEs) (Antoniou *et al.*, 2003; Benton *et al.*, 2002) derived from housekeeping genes. Both were applied in cell line development and resulted in an increased number of stable high-producers in CHO host cells.

A third approach involves improved selection strategies using preparative FACS (Mattanovich & Borth, 2006). Cell sorting can isolate transgene integration events into a rare genomic locus where high-level expression is possible and silencing does not occur (Bestor, 2000; Kito *et al.*, 2002). The cell sorting approach requires a fluorescent readout that is correlated to the expression of the recombinant product of interest. Internal ribosomal entry sites (IRES) can be used to couple the expression of the gene of interest to a marker that can be readily detected by FACS, such as GFP, or a cell surface antigen (Gaines & Wojchowski, 1999; Mancia *et al.*, 2004). Alternatively, the product itself can be captured on the cell surface by a polymer matrix or low temperature (Pichler *et al.*, 2009). Kaufman *et al.* (2008) describe the repeated isolation of the highest fluorescent cells by FACS to increase specific cellular productivity. Cell clones isolated like this provided highly homogenous and stable gene expression.

Besides FACS selection, there is the possibility to dilute and grow cells in a semi-solid methylcellulose medium until colony formation. The secreted recombinant product immobilised by the medium can be detected by fluorescence labelled antibodies as a

halo around the colony (Caron *et al.*, 2009). Cell colonies surrounded by the biggest halo were either picked up manually after microscopic identification or automatically by colony pickers resulting in a higher throughput (Caron *et al.*, 2009).

The exploitation of endogenous hot spots for transgene expression was further facilitated by targeting transgenes directly to such hot-spots instead of random integration. Site-specific recombination enables such targeting. It can improve predictability in cell line development with regard to the position effect. Moreover, time and effort to identify high and stable producers can be remarkably reduced.

1.3 Site-specific recombination systems in cell line development

1.3.1 Recombination systems

Recombination describes the physical exchange of DNA sequences based on their homology. In eukaryotes, recombination naturally occurs as a general DNA repair mechanism in mitosis or as chromosomal crossover in meiosis to provide genetic diversity. Depending on the degree of homology between the exchanged DNA sequences, homologous recombination requiring extensive sequence homology, site-specific recombination requiring short specific sequences and illegitimate recombination requiring only a few identical nucleotides are distinguished.

While homologous recombination is a popular tool for the modification of murine embryonic stem cells (Glaser *et al.*, 2005), it is hampered in differentiated cells by low frequency and is masked by more frequent illegitimate recombination (Puttini *et al.*, 2005). To enhance the frequency of homologous recombination, artificial zinc finger nucleases (ZFNs) have been evolved (Moehle *et al.*, 2007). ZFNs consist of a DNA cleavage domain fused to an engineered zinc finger binding domain. The binding domain is homologues to a known expression locus and directs the ZFN to its target. The cleavage domain stimulates the host cell's DNA repair mechanism by a site-specific double-strand break followed by the integration of donor sequences in the host cell genome using homology-directed repair mechanism (Liang & Garrard, 1999). However, this method suffers from cytotoxicity which is caused by illegitimate DNA cleavage (Porteus & Baltimore, 2003). Although an increased recognition and cleavage specificity of ZFNs is under investigation (Pruett-Miller *et al.*, 2008; Urnov *et al.*, 2005), this method has still to be developed further for the generation of production cell lines. Another challenge is the identification of appropriate chromosomal sites which are currently not

known for CHO production cell lines. In human cell lines a genomic locus was identified that allows stable long-term expression of transgenes (Smith *et al.*, 2008). It is referred to as AAVS1 (on chromosome 19, within gene PPP1R12C) and has already been used for ZFN transgenesis in different cell lines including HEK293 (DeKelver *et al.*, 2010). However, it has not been analysed yet, if this locus is also a high expression hot spot.

The technology that currently might be the best choice for targeted mammalian cell line development is based on site-specific recombination (SSR) systems of heterologous origin. SSR systems consist of the recombinase enzyme and two, double-stranded DNA recognition target (RT) sites that can recombine in a sequence dependent manner. The most common used recombinases in mammalian cell culture are the *E. coli* P1 phage-derived Cre (Cyclization recombination) (Sternberg & Cohen, 1989), the *Saccharomyces cerevisiae*-derived Flp (Flippase) (Buchholz *et al.*, 1996b) and the bacteriophage ΦC31-derived integrase (Thorpe & Smith, 1998). The corresponding recognition target sites are loxP (locus of cross-over P1) sites for Cre, FRT (Flp-recognition target) sites for Flp and a combination of attP/attB sites for ΦC31.

Both, Cre and Flp belong to the family of tyrosine recombinases, which share a common recombination mechanism based on a double-reciprocal crossover between an identical set of sites. While the sites of the two systems differ in their sequence, their overall structure is similar (Fig. 1-4). Each 34 bp long FRT and loxP site comprises an 8 bp asymmetric spacer sequence which is involved in DNA-DNA base pairing and determines the orientation of the recombination target (RT) site. In case of FRT and loxP, the spacers are flanked by two 13 bp long inverted repeats containing Flp and Cre binding sites, respectively (Abremski *et al.*, 1983; Senecoff & Cox, 1986). The natural 48 bp FRT site bears an additional base pair and another 13 bp direct repeat which is located upstream of the rest. While the third repeat is not necessarily needed for simple excision reactions, it can mediate improved recombination efficiency for other applications and serves together with the spacer region as determinant for the orientation of the FRT site.

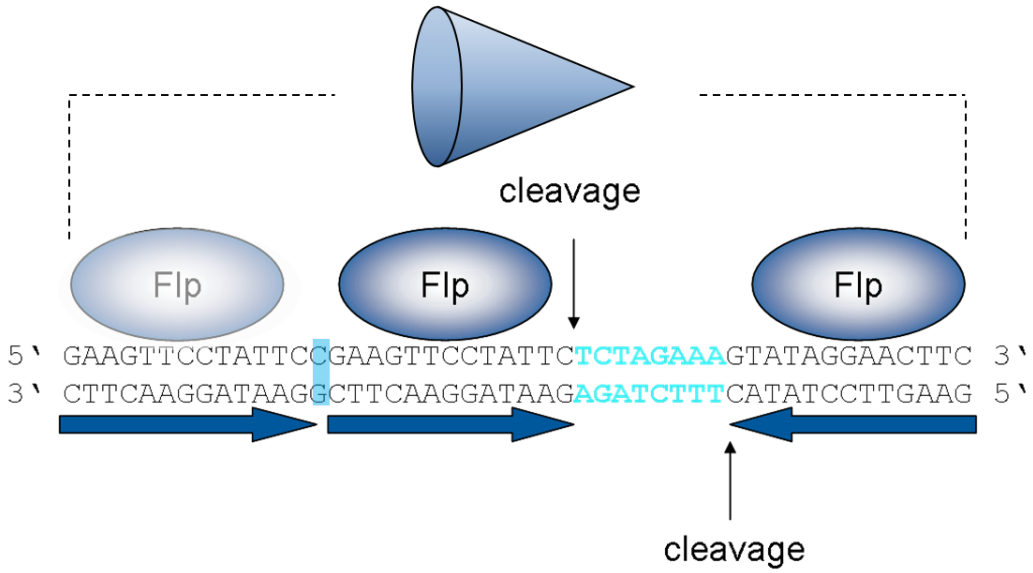


Figure 1-4: Composition of a 48 bp wild-type FRT site. DNA breaking and rejoining occurs at the asymmetric 8 bp spacer (light blue). It is flanked by three 13 bp inverted repeats representing three binding sites for Flp (horizontal arrows). Only two inverted repeats directly flanking the spacer sequence are required for a minimal FRT site with partial function (34 bp). The third repeat is located behind an isolated base pair. The vertical arrows illustrate the DNA cleavage sites of Flp. The cone represents a FRT site with its orientation.

Two RT sites are bound by four protein monomers for the recombination of two DNA sequences (Chen & Rice, 2003). Two active protomers initiate the formation of a DNA Holliday junction by a nucleophilic attack of the DNA backbone by a tyrosine. Isomerisation of the Holliday junction leads to a change in enzyme activity: the two protomers that have already catalysed a reaction are deactivated, while the other two become active. A second round of DNA cleavage and ligation reactions dissolves the Holliday junction and thereby completes the double-strand exchange. The recombination enables an excision, an insertion, an inversion or translocation of DNA sequences, depending on the position and orientation of the RT sites (Fig. 1-5). When two copies of an RT site are arranged as direct repeats, the corresponding enzyme excises the DNA sequence between the copies. The reverse reaction leads to an insertion of circular DNA containing one RT site. Due to thermodynamic reasons, the entropy-driven, monomolecular excision is favoured to the bimolecular insertion (Baer & Bode, 2001). When the RT sites are positioned in inverse orientation, the flanked DNA is inverted by the recombinase until a dynamic equilibrium is reached. Recombination of single RTs on two linear DNA molecules results in the translocation of distal sequences (Fig. 1-5).

The ΦC31 integrase, a serine recombinase, mediates recombination of two different RTs, attP and attB, resulting in the RTs attL and attR after recombination. Insertion reactions mediated by ΦC31 are unidirectional since excision would require the excisionase Xis for recombination between RTs attL and attR. Xis is not present in mammalian cells. ΦC31 recombination was commercialized for cell line development by Invitrogen (Jump-In™ system). However, there have been reported pseudo integration sites in mammalian genome for the Cre and ΦC31 mediated recombinations (Thyagarajan & Calos, 2005; Thyagarajan *et al.*, 2000). These may cause unexpected integrations cumulating in genomic instability. Pseudo integration sites have not been reported for the Flp/FRT system, but as yet there is no absolute proof.

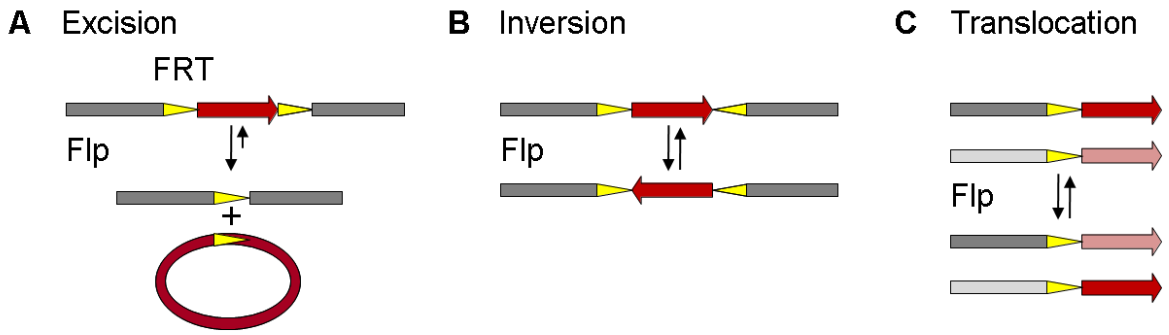


Fig. 1-5: Schematic presentation of site-specific Flp/FRT recombination with different FRT sites arrangements. (A) Flp-mediated excision of a DNA sequence that is flanked by two FRT sites with the same orientation. The reverse reaction is the integration of the circular DNA molecule. (B) Inversion of a DNA sequence that is flanked by two FRT sites with opposite orientation. (C) Translocation of DNA sequences as a consequence of recombination of two FRT sites that are located on different DNA molecules.

The wild-type Flp recombinase shows low activity at 37°C, where mammalian cells generally grow (Buchholz *et al.*, 1996b). To enhance the efficiency, mutant versions were generated. First, the temperature optimum was modified which resulted in the mutant recombinase Flpe (enhanced Flp) (Buchholz *et al.*, 1998). Based on Flpe a codon optimisation for mammalian cells resulted in the mutant FlpO (optimised Flp) and enhanced the efficiency further (Raymond & Soriano, 2007). In the past years, this development gave rise to the establishment of mammalian production cell lines by different Flp recombination strategies. Generally, these strategies combine tagging of expression hot spots with FRT sites and Flp-mediated targeting.

1.3.2 Flp-mediated excision in cell line development

A procedure exploiting Flp-mediated excision for cell line development was described by Kaufman *et al.* (Kaufman *et al.*, 2008). This approach combines the selection of highly and stably expressing GFP cell lines by preparative FACS with site-specific recombination to remove the GFP marker (Kaufman *et al.*, 2008). The stably integrated GFP gene in the tagged cells was flanked by two homospecific FRT sites (Fig. 1-6). Upon transient expression of Flp, the GFP gene was excised by Flp-mediated recombination of the FRT sites. Thereby, a downstream located gene of interest became activated by the promoter. Besides avoiding any antibiotic selection, this cell line development approach is target independent, high yields are obtainable and the cells are spared from producing GFP along with the protein of interest. However, the tagging and excision steps have to be repeated for every new cell line.

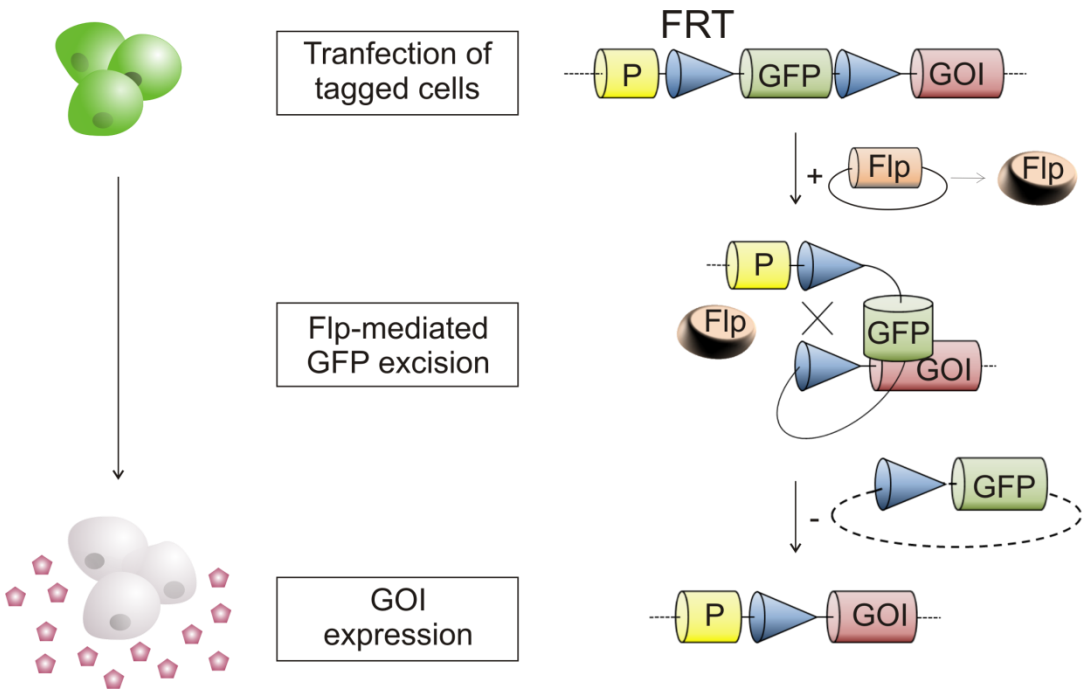


Figure 1-6: Scheme of Flp-mediated excision (FLEx) in cell line development. The pre-characterized genomic locus contained a GFP gene flanked by two FRT sites, while the promoter was located upstream of this gene cassette. The GOI was located downstream of the flanked GFP gene without an own promoter. By transfecting with a Flp expression vector, the GFP gene was excised by recombination of the FRT sites and the gene expression switched to the GOI and thus the new production cell line was established. In an authentic excision reaction the excised GFP construct got lost. Cones = FRT sites, P = promoter, pentagons = secreted protein product encoded by the GOI.

1.3.3 Flp-In™ cell lines

The Flp-In™ system commercialized by Invitrogen is based on an Flp-mediated integration strategy (Fig. 1-7). Host cells (e.g. HEK293, CHO) were established by stably integrating one copy of a tagging vector into their genome. The tagging vector contains a FRT site and a fusion gene of lacZ and zeocin resistance which enabled the selection of tagged cells by antibiotics. Site-specific recombination occurs upon co-transfection of an Flp expression vector and the Flp-In™ vector bearing the gene of interest and the required second FRT site. Since an insertion reaction is thermodynamically unfavoured compared to the excision, stringent selection systems are required (Francastel *et al.*, 1999; O'Gorman *et al.*, 1991; Schubeler *et al.*, 1998). The Flp-In™ system employs a “selection trap” relying on an incoming, non-functional, ATG-deficient hygromycin resistance gene for selection. By correctly integrating the Flp-In™ vector at the tagged locus, the non-functional hygromycin resistance gene is complemented and thus transcriptionally activated. Consequently, resistance against hygromycin is exclusively conferred to correct recombinant daughter cells while resistance against zeocin and lacZ expression is lost. Due to the transgene integration at a pre-defined locus, the resultant daughter cell clones should be isogenic and should express the gene at the same level. Thus, an extensive screening for high and stable producers is avoided. Moreover, tagged Flp-In™ cells only have to be established once and can be universally used for the integration of any GOI.

Upon Flp-In™ recombination, the entire sequence of two plasmids is present in the host genome, although most genetic elements are not essentially required, such as the fusion gene, two copies of the bacterial selection marker beta-lactamase and the replication origins. Thus, properties of the expression locus may be changed upon recombination due to a higher probability of illegitimate recombination and possible interference with the transgene (Artelt *et al.*, 1991).

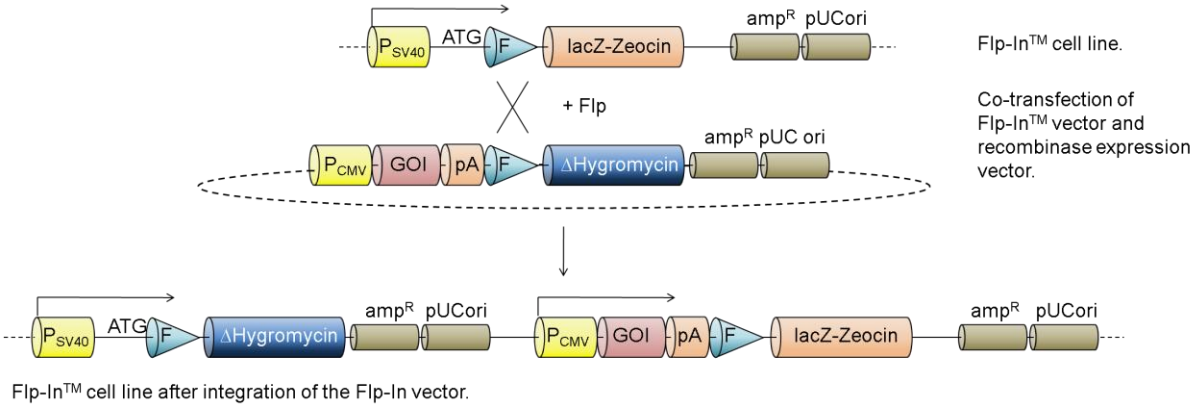


Figure 1-7: Schematic representation of the Flp-In™ system. Host cells are tagged at an expression hot spot by a simian virus 40 promoter (P_{SV40}) followed by an ATG start codon, one FRT site (F) and a lacZ-Zeocin fusion gene. Upon co-transfection of the Flp-In™ vector and a recombinase expression vector, the Flp-mediated recombination is initiated. The cells become resistant against hygromycin by targeted integration of the Flp-In™ vector due to the complementation of the ATG-deficient incoming hygromycin resistance gene. Thus, correctly exchanged cells can be selected stringently. At the same time, the GOI is integrated at the tagged expression hot spot under a cytomegalovirus promoter (P_{CMV}). ampR, PUCori = bacterial beta-lactamse gene and replication of origin, respectively. pA = poly adenylation site, GOI = gene of interest.

1.3.4 Recombinase-mediated cassette exchange in cell line development

An enhanced system for targeted cell line development is the recombinase-mediated cassette exchange (RMCE) presented by Schlake and Bode (Schlake & Bode, 1994). It circumvents the drawbacks of the Flp-In™ system and still enables the integration of user-defined DNA sequences at a tagged genomic locus (Fig. 1-8). A precondition of RMCE is the existence of heterospecific FRT sites that are recognized by Flp, but can not recombine with each other. By mutating the FRT spacer region, such FRT sites were engineered (Schlake & Bode, 1994; Senecoff & Cox, 1986; Turan *et al.*, 2010). RMCE works well with a combination of the 'F3' FRT mutant with the wild-type ('F'), since recombination of F3 sites is highly efficient, while recombination between F and F3 sites is negligible (Schlake & Bode, 1994). Other useful combinations of FRT mutants have been identified more recently (Turan *et al.*, 2010).

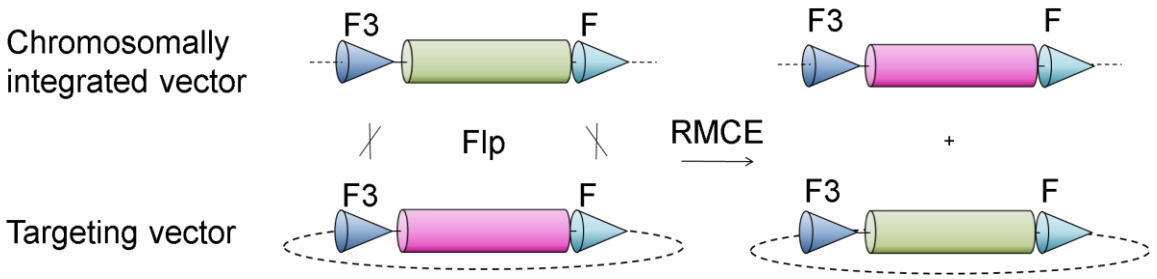


Figure 1-8: Schematic representation of RMCE with the Flp/FRT system. The exchange of DNA sequences flanked by two heterospecific FRT sites (F3 and F) is catalysed by the recombinase Flp.

Similar to other Flp recombination systems, the establishment of a RMCE system requires two steps. First, a stable and high producing genomic locus is tagged by an exchangeable gene cassette. This can be achieved either by classical antibiotic selection (Kim & Lee, 2008) or preparative FACS with a fluorescent protein as reporter (Nehlsen *et al.*, 2009; Qiao *et al.*, 2009). In the second step, the reporter is exchanged against a gene of interest by a targeted cassette exchange (Fig. 1-8). This requires that both genes are flanked by the same set of heterospecific FRT sites (e.g. F3 and F). The tagged cells are co-transfected with the targeting vector bearing the flanked gene of interest and an Flp expression vector. The transient expression of Flp catalyses a double reciprocal crossover of compatible FRT sites whereby the DNA sequences between the FRT sites are exchanged. The transfection vectors and the Flp recombinase are lost after some time. Pre-defined chromosomal loci were reused and extensive screening procedures were circumvented by RMCE (Baer & Bode, 2001; Bode *et al.*, 2000; Oumard *et al.*, 2006). Thus, cell line development clearly profits from RMCE by high expression levels and predictable cultivation properties as well as a shortened development timeline.

Stringent selection methods optimise the efficiency of cassette exchange and counter-select random integration of the targeting vector. One strategy is the selection trap. Verhoeyen *et al.* (Verhoeyen *et al.*, 2001) cloned a non functional neomycin phosphotransferase gene (Δneo^R) lacking the start codon directly downstream of the FRT cassette. Upon recombination, a promoter and a start codon, located on the targeting vector, transcriptionally activated the incomplete Δneo^R gene at the tagged locus (Fig. 1-9). The selection trap allows selection for correct RMCE events with the antibiotic G418. Alternatively, a negative selection marker such as the thymidine kinase gene may be included in the initial tagging construct, thus allowing the specific

elimination of all cells without exchanged cassette by ganciclovir selection (Toledo *et al.*, 2006; Wong *et al.*, 2005).

The promoter trap describes a combination of a FRT-cassette downstream of an activating promoter and a promoterless targeting vector. Thus, only RMCE, but not random integration, can lead to expression of the gene of interest on the targeting vector (Cobellis *et al.*, 2005).

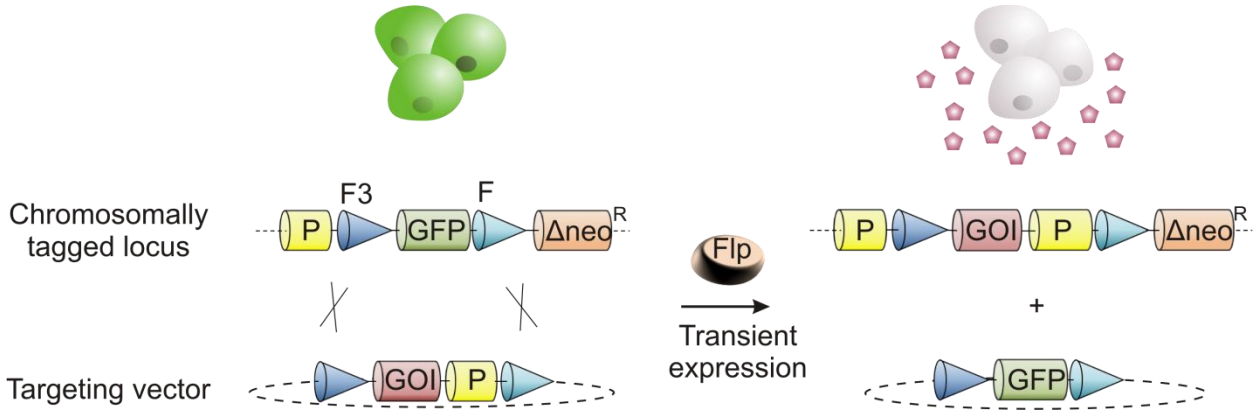


Figure 1-9: Schematic presentation of RMCE with a neomycin selection trap and a promoter trap. The scheme summarizes the strategy for selection of successfully targeted clones that was applied in the present work. The neomycin selection trap is based on the ATG-deficient neomycin resistance gene (Δneo^R) in the tagged locus. Thus, parental cells are sensitive to G418. The missing promoter (or internal ribosomal entry site, IRES) and the start codon are located on the incoming targeting vector. Upon RMCE, these elements are positioned upstream of the neomycin resistance gene, which becomes activated and confers resistance against G418. The promoter trap requires a promoter upstream of the exchangeable gene cassette. The incoming gene does not need an additional promoter. Thus, GOI expression indicates a correct recombination. Cones = FRT sites (F3 and F), P = promoter, pentagons = secreted protein product encoded by the GOI.

Since its development, RMCE has been applied in different ways for generating cell lines producing viruses (Schucht *et al.*, 2006), erythropoietin (Kim & Lee, 2008) or antibodies (Nehlsen *et al.*, 2009). In this work, it was used to establish cell lines for the production of mammalian glycoproteins for structural analysis by crystallography.

1.4 Mammalian proteins for structural analysis

1.4.1 Regulation of recombinant protein production

Recombinant protein production in mammalian cells is a complex process requiring transcription, mRNA processing and transport, translation, post-translational modification, protein folding and subcellular transport through the secretory pathway.

Although transcription is traditionally considered to be the most dominant factor controlling expression level and stability, all steps are tightly regulated in the host cell line and have an impact on the specific productivity of any producer cell line (Barnes *et al.*, 2003). Strategies to augment productivity aim, for instance, at the architecture of the expression cassette. The choice of promoter, enhancer, poly-adenylation signal, as well as the order, orientation and distances between these elements contribute to the overall expression strength (Kim *et al.*, 2009). Additionally, the sequence of the transgene itself can be optimised for higher production by adjusting codon usage and optimizing RNA secondary structure and stability (Kalwy *et al.*, 2006). For secreted proteins, it has been recognized that the secretion capacity of a cell can become limiting in the presence of high recombinant mRNA levels (Peng & Fussenegger, 2009). For secretion, newly synthesized proteins pass through a series of compartments within the cell, including the rough endoplasmatic reticulum (rough ER) and the Golgi apparatus (Palade, 1975). Genetic engineering was applied at different points to enhance the secretion of recombinant proteins. First approaches altered the ER environment, an important region for glycosylation, folding and assembly of proteins. Overexpression of molecular chaperones assisting the folding of newly synthesized proteins was induced. However, overexpression of the chaperon BiP/GRP78 and the protein disulfide isomerase (PDI) led to contradictory results for different proteins. While PDI overexpression in CHO cells reduced the expression of a TNFR:FC fusion protein (Davis *et al.*, 2000), the specific production rate of an antibody was increased by 40% (Borth *et al.*, 2005). Further engineering approaches successfully focused on the overexpression of ceramide transport protein (CERT) which is involved in protein transport from Golgi apparatus to the plasma membrane (Florin *et al.*, 2009) or the X-box binding protein 1 (XBP-1) which is a transcription activator affecting multiple key genes in the secretory pathway (Becker *et al.*, 2008). However, it was demonstrated that XBP-1 expression enhanced apoptosis in CHO cells (Becker *et al.*, 2010). A combinatorial approach co-expressing XBP-1 and the anti-apoptotic protein X-linked inhibitor of apoptosis (XIAP) achieved elevated specific productivities and final titres in fed-batch cultures of polyclonal cell lines when compared with control cells.

Anti-apoptose engineering strategies are considered to be an important factor for viable cell density and thus protein yield and have already been applied before (Mastrangelo *et al.*, 2000). Most strategies focus on the overexpression of anti-apoptosis genes of the

Bcl-2 family (Lee & Lee, 2003) or caspase inhibitors such as XIAP (Sauerwald *et al.*, 2002).

1.4.2 Glycoprotein production for X-ray structure analysis

Beyond their production in sufficient amounts, glycoproteins pose another challenge for X-ray structure analysis. Though being required for folding and transport through the secretory pathway via endoplasmatic reticulum and the Golgi apparatus, the flexible and heterogeneous nature of N- and O-linked mammalian glycosylation generally hampers crystallisation (Butters *et al.*, 1999). The glycans may also mask possible crystal contacts on the protein surface. Heterogeneity of the carbohydrate chains is determined by variable glycan composition (micro-heterogeneity) and non-uniformly occupied sites (macro-heterogeneity). Since O-linked glycans are often located in extended, unfolded, serine-, threonine- and proline-rich regions of a protein, they can generally be removed by mutagenesis or generating expression constructs without the corresponding domain (Nettleship *et al.*, 2010). On the contrary, N-linked glycosylation of the Asn-X-Ser/Thr motif is generally necessary for protein folding and stability. Variably occupied sites are regarded as not essential for folding and stability and thus can be removed by genetic engineering (Nettleship *et al.*, 2007). The composition of N-linked glycosylation can be classified in high-mannose, complex and hybrid type glycans of different length and branch number (Fig. 1-10 A). To circumvent the problems due to essential glycans, glycoproteins can be produced in the presence of glycosylation pathway inhibitors or directly by glycosylation mutant cell lines. The addition of glycosylation pathway inhibitors like N-butyldeoxynojirimycin (NB-DNJ), swainsonine and kifunensine to a production cell line results in glycoproteins with truncated carbohydrate chains (Fig. 1-10 B) (Chang *et al.*, 2007). Glycans of high-mannose type can be further truncated efficiently to a single N-acetylglucosamine (GlcNAc) by endoglycosidase H (endo H). Although all simplified glycans are sensitive to endo H cleavage, the truncation is most efficient with kifunensine treated glycoproteins bearing $Man_9GlcNAc_2$ sugar chains (Nettleship *et al.*, 2010). Thus, for crystallisation projects, kifunensine has been used with endo H treatment (Bishop *et al.*, 2009; Bowden *et al.*, 2008a; Bowden *et al.*, 2008b; Carafoli *et al.*, 2009), but also without (Crispin *et al.*, 2009). The inhibitory drugs can be combined with protein production in glycosylation mutant cell lines for an even higher stringency of glycan truncation. Various mutant CHO and HEK293 cell lines are available synthesizing glycoproteins with truncated carbohydrates. CHO Lec3.2.8.1 and

HEK293S cells, for instance, do not process N-linked glycans beyond the high-mannose type $Man_5GlcNAc_2$ (Fig. 1-10 C) (Reeves *et al.*, 2002; Stanley, 1989). Their protein products are therefore more homogenous and more likely to crystallize (Davis *et al.*, 1993). For a further truncation, the N-linked glycans are amenable to endo H cleavage (Chang *et al.*, 2007; Wilke *et al.*, 2010). The use of mutant cell lines enabled the crystallization of various glycoproteins (Niemann *et al.*, 2007; Reeves *et al.*, 2002).

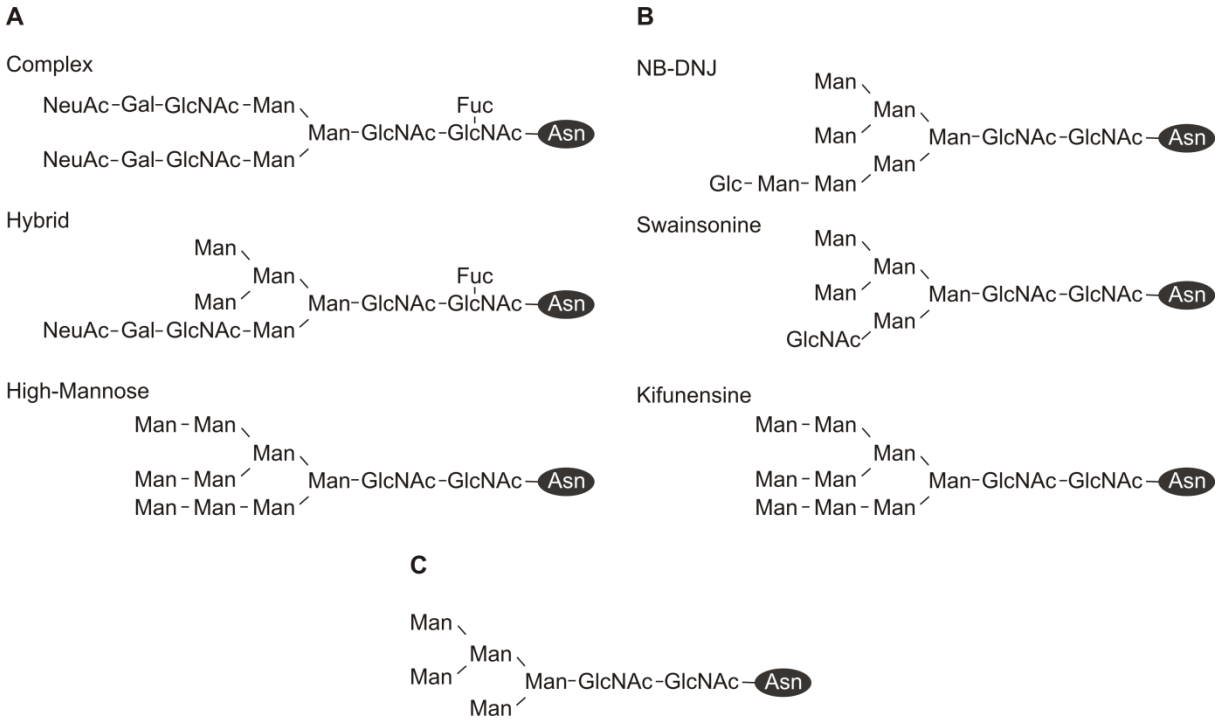


Fig. 1-10: Diagramatic representation of the major N-linked glycans produced by (A) mammalian cells, (B) CHO and HEK293 cells in the presence of the N-glycan processing inhibitors N-butyldeoxynojirimycin (NB-DNJ), swainosine and kifunensine and (C) the mammalian mutant cell lines CHO Lec3.2.8.1 and HEK293S.

1.4.3 Hepatocyte growth factor

Human hepatocyte growth factor (HGF) acts as a mitogen in cultured primary hepatocytes (Nakamura *et al.*, 1984; Nakamura *et al.*, 1989). It was independently discovered and described as strong motility or scatter factor (SF) for epithelial cells (Stoker & Perryman, 1985). The identity of HGF and SF was demonstrated later (Weidner *et al.*, 1991). By binding to the receptor tyrosine kinase c-Met, HGF/SF activates intracellular signalling pathways resulting in versatile biological responses (Fig. 1-11). These are involved in promoting cell proliferation, survival, migration, neurite extension, angiogenesis, wound repair and cell migration during development (Birchmeier *et al.*, 2003; Chmielowiec *et al.*, 2007; Funakoshi & Nakamura, 2003; Huh

et al., 2004). Deregulated activation of the HGF/SF receptor c-Met is associated with tumour invasion in many cancers with an aggressive phenotype and poor prognosis (Birchmeier *et al.*, 2003; Eder *et al.*, 2009). Thus, inhibition of Met activity, either by small molecule kinase inhibitors or by protein-based HGF/SF antagonists, has become an important strategy for developing anticancer therapeutics (Tolbert *et al.*, 2010).

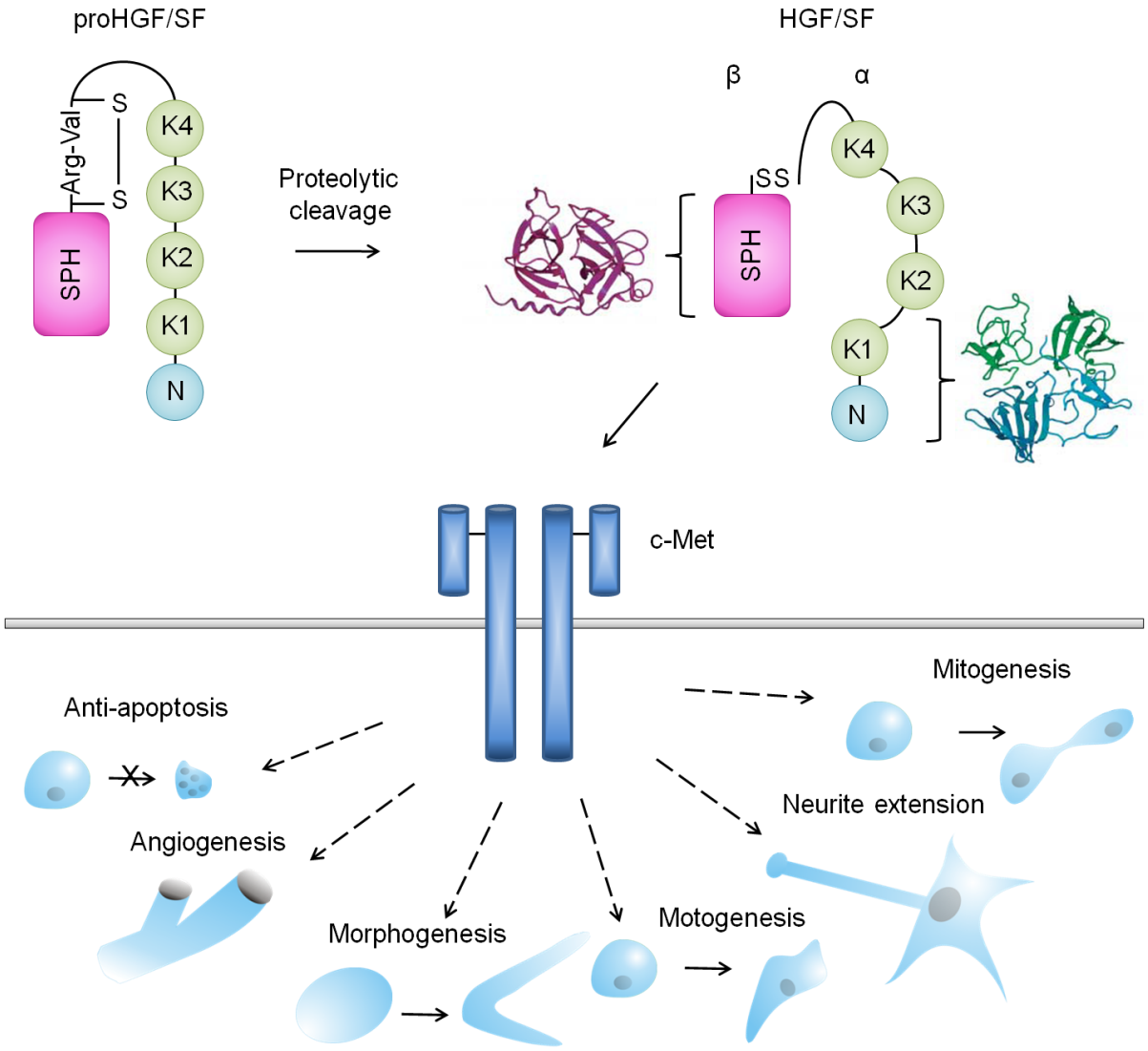


Figure 1-11: Schematic representation of HGF/SF and cellular responses of c-Met signalling upon stimulation with HGF/SF. Mature HGF/SF is generated by proteolytic cleavage from its single-chain precursor pro-HGF. By binding to the receptor tyrosine kinase c-Met, HGF/SF activates various intracellular signalling pathways. N: N-terminal domain; K1-4: Kringle domains; SPH: serine protease homology domain. (Crystal structures of the NK1 and SPH domains derive from Birchmeier *et al.*, 2003).

Structurally, human HGF/SF is a disulphide-linked heterodimer containing 698 residues and four N-linked glycosylation sites. The 57.3 kDa α-chain comprises an N-terminal loop domain and four kringle domains, while the 22.6 kDa β-chain contains a serine

protease like domain without enzymatic activity (Fig. 1-11) (Miyazawa *et al.*, 1989). Despite the missing catalytic function, HGF/SF belongs to the family of serine protease growth factors. It is synthesised and secreted as a biological inactive single chain precursor (proHGF) which binds to Met, but is not capable to activate it (Gherardi *et al.*, 2006). A serine protease mediated cleavage at arginine-494 induces conformational changes leading to an active two chain form of HGF/SF (Stamos *et al.*, 2004). Insights in this activation mechanism have been exploited by generating synthetic versions of proHGF that can not be activated. While one approach aimed at modifying the "activation pocket" of the HGF/SF β-chain (Kirchhofer *et al.*, 2007), Gherardi *et al.* (Gherardi *et al.*, 2006) mutated arginine-494 in the recognition site to glutamate which resulted in a non cleavable single chain variant of HGF/SF (scHGF). A detailed three dimensional structural analysis of the respective scHGF variant in complex with the c-Met receptor may give further insights in conformational changes of HGF/SF upon cleavage and the activation process of receptor tyrosine kinases.

1.4.4 LAMP-3

The human lysosome associated membrane protein 3 (LAMP-3) comprises 416 amino acids and belongs to the family of lysosome associated glycoproteins. It is also known as DC-LAMP, since its mRNA is exclusively detected in dendritic cells (DC) in lymphoid organs (de Saint-Vis *et al.*, 1998) and in lung type II pneumocytes (Akasaki *et al.*, 2004). Its expression increases steadily during DC maturation and sharply and strongly upon DC activation by lipopolysaccharide, tumour necrose factor α or CD40L. Shortly before the MHC class II molecules are translocated to the cell membrane, LAMP-3 is detectable in the MHC class II compartment. Due to its localisation it is assumed that LAMP-3 function in lysosomes is related to presentation of peptide-MHCII complexes at the DC surface (de Saint-Vis *et al.*, 1998). Overexpression of LAMP-3 in tumour cells was found to coincide with an increased potential of metastasis and thus may serve as a prognostic factor for diverse cancer types (Kanao *et al.*, 2005; Ozaki *et al.*, 1998).

Members of the LAMP family are type I transmembrane proteins and thus have one transmembrane helix (Fig. 1-12). The helix connects the short cytoplasmic part containing a glycine-tyrosine motif with a luminal part, consisting of two domains. LAMP-3 is homologous to other family members such as LAMP-1 (28.9% sequence identity), LAMP-2 (26.9% sequence identity) and CD68/macrosialin (28.1% sequence identity). LAMP-1 and LAMP-2 are involved in lysosomal biogenesis, lysosomal fusion

of phagosomes and autophagy (Eskelinen & Saftig, 2009; Huynh *et al.*, 2007). The two luminal domains of LAMP-1 and 2 contain four disulphide bonds each. LAMP-3 and human CD68/macrosialin (a dendritic cell/macrophage marker) differ from the other LAMP members, as they only contain four cysteines that may form disulphide bonds. These are located in the membrane proximal half of the luminal region. The membrane distal domains contain several serine, threonine and proline residues that are thought to be modified by O-linked glycans as it is the case in mucin-like proteins (Holness & Simmons, 1993). The LAMP-3 amino acid sequence also contains seven potential N-linked glycosylation sites. Thus, the observed molecular mass in SDS-PAGE is around 70-90 kDa, whereas the calculated mass for the polypeptide core is 44 kDa (de Saint-Vis *et al.*, 1998). Knowledge of the structure of the luminal LAMP domain is currently limited to the position of N-glycosylation sites and the disulphide bridge configuration. Structural analysis of LAMP-3 luminal domains would reveal its unknown three-dimensional structure which could provide information about its molecular mode of action.

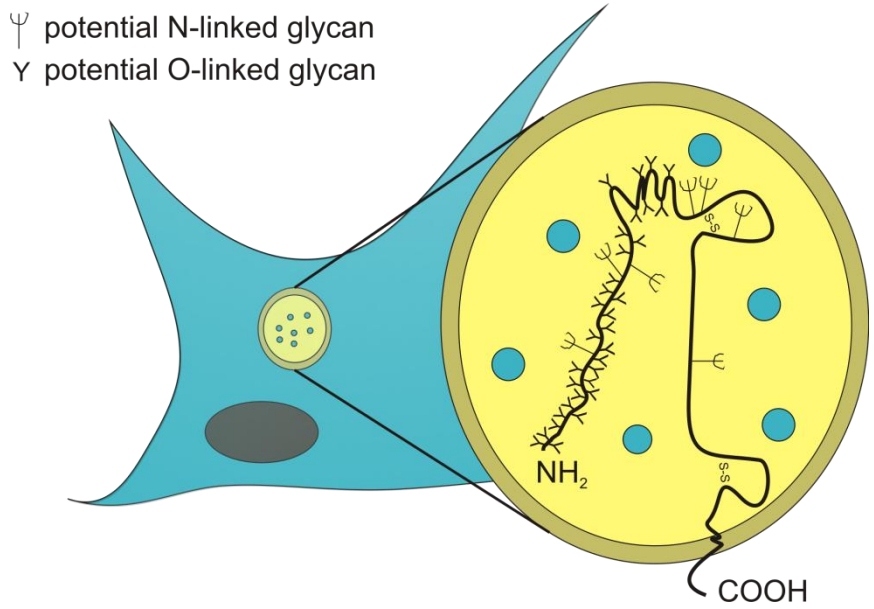


Figure 1-12: Schematic presentation of LAMP-3. LAMP-3 is localized in lysosomes of dendritic cells. N- and O-glycans are indicated by tridents and Ys, respectively. The position of the disulphide bonds is indicated. The larger part of the molecule resides in the lysosomal lumen.

1.5 Aims of the thesis

Structure determination of mammalian glycoproteins often calls for recombinant production in mammalian cells due to their complex processing. X-ray structure analysis

frequently requires generating several genetic constructs per protein to create well crystallisable material. However, establishing stable production cell lines remains time-consuming and expensive. Additionally, the heterogeneous and complex glycosylation of mammalians complicates crystallisation of glycoproteins. Thus, strategies capable of dealing with the glycosylation problem and enabling straightforward cell line development are of essential interest.

The primary aim of this work was the development of strategies simplifying the production and crystallization of glycoproteins. The glycosylation mutant CHO Lec3.2.8.1 cell line that produces more homogenous glycoprotein suitable for crystallisation was selected as host cell. The CHO Lec3.2.8.1 genome should be modified using Flp-mediated recombination in order to optimise cell line development towards higher-throughput and predictable expression strength. To this end, two strategies were available for evaluation, namely a GFP excision model (Kaufman *et al.*, 2008) and RMCE (Schlake & Bode, 1994; Schucht *et al.*, 2006). Possible target proteins were RFP, HygTk, human scHGF and a human luminal membrane-proximal domain of LAMP-3 (hLAMP3-prox).

At the beginning of this work no member of the LAMP family had been structurally characterized and no homologous structure had been available for phase determination by molecular replacement. Thus, another aim of this thesis was the determination of the crystal structure of hLAMP3-prox domain as a representative for LAMP family members. Upon cell line development, procedures for the recombinant production, purification, crystallisation and crystal derivatisation were to be established.

2 Material and Methods

2.1 Instruments

Table 2-1: Instruments

Instrument	Model	Company
Autoclave	Steam-Sterilizer 9-6-9	Belimed
Biochemical	YSI 2700 select	Yellow Springs Instruments
Analyser	Trans-Blot SD	BioRad
Blot chamber	CASY Model TTC	Innovartis AG
Cell counter	BB15	Heraeus
Cell culture	Forma Reach-In CO_2	Thermo Scientific
incubator	Biofuge pico	Heraeus
	Multifuge 1_{S-R}	Heraeus
Centrifuges	5417R	Eppendorf
	4K15C	Sigma
	RC6	Sorvall Instruments
	RC12BP	Sorvall Instruments
Centrifuge rotor	SA-600, SLA-3000	Sorvall Instruments
Clean Benches	Hera Safe KS15	Thermo Electron Corporation
Cryostorage system	10 K series cryostorage system	Taylor-Wharton
Gel electrophoresis	Wide Mini Sub Cell GT	BioRad
chambers	Sub-cell Model 96	BioRad
Electroporator	GenePulser Nucleofector®	Amaxa
FACS	MoFlo	Beckman Coulter
	FACSAria	Becton Dickinson
Flow cytometer	Easy Cyte Mini	Guava Technologies
	LSR II	Becton Dickinson
Fluorescence	Axiovert 1000	Zeiss
microscope	Digital camera: DMX1200	Nikon
FPLC system	Äkta-FPLC (UPC-900, P-920)	Amersham pharmacia biotech
Incubator	B20	Heraeus
Incubation shaker	Multitron II	Infors
Laboratory shaker	WT12	Biometra
Light microscope	CKX4	Olympus
	SZ40	Olympus
	Digital camera: DP11	Olympus
Luminometer	LAS 3000	FujiFilm
pH meter	S20-K Seven Easy pH	Mettler Toledo
Photometer	NanoDrop ND-1000	PeqLab Biotechnology
Pipetting robot	Mosquito	TTP Labtech
Power supply	Power-Pac 300	BioRad
SDS-PAGE system	Mini Protean III	BioRad
Speed Vac	RC 1010	Thermo Electron Corporation
Thermo cycler	T-Gradient	Biometra
Thermo mixer	5436	Eppendorf
UV-imaging system	Gel Logic212	Kodak

Instrument	Model	Company
Vortexer	Vortex Genie2™	Scientific Industries
Water purifier	Milli-Q Plus	Millipore
X-ray generator	RU-H3R	Rigaku MSC

2.2 Material

2.2.1 Chemicals and cell culture material

If not stated otherwise, chemicals were purchased from the following companies: Amersham, Bayer, Biolabs, Boehringer, Clontech, Difco, Fluka, Gibco, GE-Healthcare, Hampton Research, Hoechst, Invitrogen, Merck, Millipore, MoBiTec, Novagen, Promega, Qiagen, Riedel de Haen, Roche, Roth, Serva, Sigma-Aldrich and Stratagene. The quality standard was “pro analysis” (p.a.). Polystyrene, polycarbonat and glass cell culture materials were procured by Falcon, Greiner and Nunc.

2.2.2 Enzymes

Table 2-2: Enzymes

Name	Company
Benzonase	Novagen
CIP (Calf Intestine Phosphatase)	New England Biolabs (NEB)
Expand High Fidelity^{Plus} Polymerase	Roche
Restriction endonucleases	New England Biolabs (NEB)
RNase A	Boehringer Mannheim
T4 DNA-Ligase	New England Biolabs (NEB)
T4 DNA-Polymerase	New England Biolabs (NEB)
Taq Polymerase	New England Biolabs (NEB)
Trypsin/ EDTA	PAA Laboratories

2.2.3 Molecular weight standards

Table 2-3: Molecular weight standards.

Name	Type	Usage	Company
DNA Molecular Weight Marker II, DIG-labeled	DNA	Southern Blot	Roche
λ *Bst*P I digest	DNA	Agarose gel electrophorese	Takara
Precision Plus Protein Unstained	Protein	SDS-PAGE	BioRad
Precision Plus Protein All Blue	Protein	SDS-PAGE/Western Blot	BioRad

2.2.4 Media

E.coli medium

LB-medium (Luria-Bertani):	Tryptone (Bacto)	10 g/L
	Yeast extract (Bacto)	5 g/L
	NaCl	5 g/L
LB-agar:	LB-medium +	
	agar (Bacto)	16 g/L
SOC-medium:	Tryptone (Bacto)	20 g/L
	Yeast extract (Bacto)	5 g/L
	NaCl	10 mM
	KCl	2.5 mM
	Glucose	2 mM
	MgCl2	10 mM
	MgSO4	10 mM

If not stated otherwise, media for bacterial cultures were autoclaved 20 min at 121°C. Heat instable components such as glucose and antibiotics were added after filtration (0.2 µm).

Mammalian cell medium

ProCHO5-medium:	Lonza; supplemented with 7.5 mM L-glutamine (Roth) and 11 mg/L phenolred (Sigma)	
ProCHO5 + ZKT-I [1:1]: medium	ProCHO5 powder (Cambrex)	13.8 g/L
	ZKT-I powder (Biochrom)	6.1 g/L
	NaHCO3 (Merck)	2.9 g/L
	L-glutamin (Roth)	3.8 mM
	NaOH	2 mM
	Human Insulin (Cambrex)	0,1% (v/v)
	2x CHO7 lipids supplement (Cambrex)	0.04% (v/v)
CD-Hybridoma medium:	Invitrogen; supplemented with 8 mM L-glutamine (Roth) and 5-10% FCS (Gibco)	

If not stated otherwise, supplements were filtered (0.2 µm) before addition to the medium. Powder media for mammalian cell cultures were sterilized after solving in purified water (MilliQ) by a filtration line (Pall) combining two nylon filter units with decreasing pore size (1 µm, 0.2 µm).

2.2.5 Antibiotics

The following antibiotics were used for various cultured cells to ensure the selectivity of liquid and solid media.

Table 2-4: Antibiotics.

Name	Stocking concentration	Final concentration
Ampicillin	100 mg/ ml	100 µg/ ml
G418	100 mg/ ml	1.5 mg/ ml
Hygromycin B	100 mg/ ml	150 µg/ ml

2.2.6 Bacterial strains

Table 2-5: Bacterial strains.

E.coli strain	Genotype	Source
Top 10	F-, *mcr* A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80 *lac*Z Δ M15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal* U *gal* K *rps*L *end*A1*nup*G.	Invitrogen
XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacIqZΔM15 Tn10 (Tetr)].*	Stratagene
DH5α	*F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZ* ΔM15 Δ*(lacZYA- argF)*U*169, hsdR17(rK- mK +), λ-.*	Invitrogen

2.2.7 Mammalian cell line

CHO Lec3.2.8.1 This line is a glycosylation mutant clone derived from the parental Chinese hamster ovary (CHO) clone Pro-5 (a proline auxotroph, ATCC no. CRL 1781) by selection for resistance to different lectins (Stanley, 1989). This clone was adapted to grow in suspension.

2.2.8 Oligonucleotides

Oligonucleotides were used as primers for polymerase chain reactions (PCRs) and DNA sequencing. They were purchased from MWG Eurofins Operon (desalted or HPSF purified).

Table 2-6: Oligonucleotides.

No.	Name	Length (bp)	Sequence (5' → 3')
1	AY188393_F	17	AAGCAGCTGGCGCTGGA
2	AY188393_R	17	TGGAGGCCACCTCCCGA
3	BGH_rev	18	TAGAAGGCACAGTCGAGG
4	CassettepEF_S	28	TCGGGAGATCTCGACCGAGCTTTGCAAA

No.	Name	Length (bp)	Sequence (5' → 3')
5	CassettepEF_AS	30	TGCTCGAGCGGCCGCTCTAGAACTAGTGGA
6	CMVfor	21	CGCAAATGGGCGGTAGGCGTG
7	dneo1-AS	24	CAGTTCATTCAGGGCACCGGACAG
8	dneo-pEF-FS-2	37	TAGATGCATGCTCGAGCGACTCTAGAGGATCCCC CGA
9	dneo-pEF-S-1	36	AGGAACTTCGGAATTCAGTGGATTGCACGCAGGT TC
10	flpo_as	24	GGAGGGAGAGGGGCGGAATTCTTA
11	flpo_s	25	CGGCGGCTCTAGAATGGCTCCTAAG
12	GFPKassette_AS	29	GCAGCCAGGGCGTGGAAGTAATTCAAGG
13	GFPKassette_S	29	TCACATGGTCCTGCTGGAGTTCGTGACCG
14	HygTk_771S	19	CGAGGTCGCCAACATCTTC
15	HygTk_HpaI_as	30	GTATCCAATTGCGATTAGCCTCCCCCATCT
16	HygTk_IRES_s	30	ACACGATAAGCTATGAAAAAGCCTGAACTC
17	HygTk-pEF-1	36	CCGGTCGCCACCATGAAAAAGCCTGAACTCACCG CG
18	HygTk-pEF-2	39	ATCGCCAGTCACGCGTCAGTTAGCCTCCCCCATC TCCCG
19	IRES_as	38	GAGTTCAGGCTTTTTCATAGCTTATCATCGTGTTT TTC
20	IRES_s	27	ACTACCAATTGCAGGTAAGTATCAAGG
21	IVS_as	18	TGAACTGGGAGTGGACAC
22	KonkatemerGFP Kassette_AS	29	GCAGCCAGGGGCGTGGAAGTAATTCAAGG
23	KonkatemerGFP Kassette_S	29	TCACATGGTCCTGCTGGAGTTCGTGACCG
24	KonkatemerRFP Kassette_AS	28	CGGAGGAGGCCATGGTGGCGACCGGAAA
25	KonkatemerRFP Kassette_S	30	CTGCCCGCGCGGTGTTCCGCATTCTGCAAG
26	KonkatemerTest_ asPr	30	TCGAGCGGCCGCTCTAGAACTAGTGGATCC
27	KonkatemerTest_ s	20	AAAGGAGTGGGAATTGGCTC
28	pEF-F3/2-GFP-MCS-for	18	TCACTGCATTCTAGTTGT
29	pEF-FS-AS2	25	GAATTCCGAAGTTCCTATACTTTCT
30	pEF-FS-S	28	GTGTCGTGAGGAATTAGCTTGACCAAGC
31	pEF_qPCR_F	17	CCCGAGGGTGGGGGAGA
32	pFS-S	30	GGAGATCAGCTTGAAGTTCCTATTCCGAAG
33	PGK_AS	28	CCTGGGGAGAGAGGTCGGTGATTCGGTC
34	pgkatg-pfsrfp-1	65	GTTCCTGTAGCGGCCGCACGCGTATTTAAATGAT CCAGACATGATAAGATACATTGATGAGTTTG
35	pgkatg-pfsrfp-2	79	TAGATGCATGCTCGAGAAGTTCCTATACTTTCTAG AGAATAGGAACTTCGGAATAGGAACTTCTCCATGT TGGCGACCG
36	SEAP_AS	40	CGATCAATTGAGATCTCTCAACCCGGGTGCGCGG CGTCGG
37	SEAP_S	50	CGATCTTCGAATTTCCGATCGATCCACCATGCTGC TGCTGCTGCTGCTGC

2.2.9 Vectors

The following vectors were used.

C53	pBlueScript vector bearing scHGF cDNA. (Gherardi, E., MRC, Cambridge)
$pEF_{F3}EGFP_{F3}mcs$	Tagging vector of the GFP-excision system. The EGFP gene is flanked by two F3 FRT sites. A multiple cloning site is located downstream of the second FRT site. The gene cassette is under control of the human elongation factor 1-α (EF1-α) promoter (Kaufman *et al.,* 2008).
$pEF_{F3}EGFP_{F3}scHGF$	Tagging vector of the GFP-excision system to generate scHGF production cell lines. The scHGF gene including a C-terminal His_6-tag and a mouse immunoglobulin signal peptide (Swissprot P01750) were cut out from C53 by XbaI and cloned into $pEF_{F3}EGFP_{F3}mcs$ cut by SpeI.
$pEF_{F3}EGFP_{F3}LAMP3$	Tagging vector of the GFP-excision system to generate hLAMP3-prox production cell lines. A sequence encoding a mouse immunoglobulin signal peptide (Swissprot P01750), amino acids 222-381 (VKTG...SDYT) of LAMP-3 (GenBank AAH32940) and a His_6-tag was cloned between the BamHI and NotI sites of $pEF_{F3}EGFP_{F3}mcs$.
pEF-FS-EGFP	Tagging vector of a RMCE system expressing EGFP under control of the human EF1-α promoter. EGFP and a SV40 polyA site are flanked by F3 and wild type FRT sites. (Gossen, M., MDC, Berlin)

pEF-FS-EGFP-dneo	EGFP expressing tagging vector of a RMCE system with neomycin selection trap. An ATG-deleted neomycin phosphotransferase (Δneo^R) gene was PCR amplified with primers 8 and 9 on ptagSVHTGdneodC and cloned between EcoRI and XhoI in pEF-FS-EGFP by the In-Fusion Cloning system (Clontech).
pF3LMCSF_RFP_Express	Eukaryotic expression vector of DsRed Express under control of a CMV IE promoter. (Qiao, J., HZI, Braunschweig)
pFS-EGFP	EGFP targeting vector compatible with the pEF-FS-EGFP RMCE system. The vector pEF-FS-EGFP was cut by BglII and HindIII, blunted and ligated.
pFS-EGFP-IRES-HygTk	EGFP and HygTk targeting vector compatible with the pEF-FS-EGFP RMCE system. The IRES element was PCR amplified with primers 19/ 20 on pIRESneo3 (Clontech) while the HygTk gene was PCR amplified with primers 15/ 16 on ptagSVHTGdneodC. The primers were designed to mediate a homologous recombination in a third PCR amplification (primers 15/ 20) resulting in a 2.95 kb fragment. The third PCR product and the vector pFS-EGFP were cut by MfeI/ HpaI. The vector was dephosphorylated and ligated with the PCR product.
pFS-GILT-PGK	GILT targeting vector compatible with the pEF-FS-EGFP-dneo RMCE system. (Büssow, K., HZI, Braunschweig)

pFS-HygTk	HygTk targeting vector compatible with the pEF-FS-EGFP RMCE system. After EGFP excision in pFS-EGFP by NcoI and MluI the PCR amplified HygTk gene from ptagSVHTGdneodC (primers 17/ 18) was inserted by the In-Fusion cloning system (Clontech).
pFS-scHGF-PGK	scHGF targeting vector compatible with the pEF-FS-EGFP-dneo RMCE system. The vector pFS-RFP-PGK was cut by BstBI, blunted, cut by NotI and dephosphorylated. It was ligated with a 2.2 kb fragment of $pEF_{F3}EGFP_{F3}$scHGF cut by SmaI and NotI.
pFS-scHGF-IRES-HygTk	scHGF and HygTk targeting vector compatible with the pEF-FS-EGFP RMCE system. (Duda, A., HZI, Braunschweig)
pFS-FcεIα-PGK	FcεIα targeting vector compatible with the pEF-FS-EGFP-dneo RMCE system. (Owens, R., OPPF, Oxford)
pFS-hLAMP2lum-PGK	hLAMP2-lum targeting vector compatible with the pEF-FS-EGFP-dneo RMCE system. (Büssow, K., HZI, Braunschweig)
pFS-hLAMP3-PGK	hLAMP3-prox targeting vector compatible with the pEF-FS-EGFP-dneo RMCE system. (Büssow, K., HZI, Braunschweig)
pFS-mLAMP2lum-PGK	mLAMP2lum targeting vector compatible with the pEF-FS-EGFP-dneo RMCE system. (Büssow, K., HZI, Braunschweig)

pFS-NALP3-H8-PGK	NALP3-His_8 targeting vector compatible with the pEF-FS-EGFP-dneo RMCE system. (Polle, L., HZI, Braunschweig)
pFS-RFP	RFP targeting vector compatible with the pEF-FS-EGFP RMCE system. The vector pF3LMCSF_RFP_Express was cut by NotI. The 1.55 kb fragment was made blunt-end and cut with NcoI. The 0.63 kb fragment was ligated into the 2.5 kb fragment of a MluI/ NcoI cut pFS-EGFP vector.
pFS-RFP-PGK	RFP targeting vector compatible with the pEF-FS-EGFP-dneo RMCE system. The PGK promoter sequence and the ATG start codon were PCR amplified on pPGKMFGeGFP1. The primers 34/ 35 were designed according to the In-Fusion PCR protocol (Clontech). An additional 5' tail of the antisense primer contained a wild type FRT sequence. The 1022 bp PCR product was cloned into the NotI/ XhoI cut vector pFS-RFP by the In-Fusion PCR system.
pFS-rLAMP2dist-PGK	rLAMP2-dist targeting vector compatible with the pEF-FS-EGFP-dneo RMCE system. (Büssow, K., HZI, Braunschweig)
pFS-rLAMPlum-PGK	rLAMP2-lum targeting vector compatible with the pEF-FS-EGFP-dneo RMCE system. (Büssow, K., HZI, Braunschweig)
pFS-SEAP-IRES-HygTk	SEAP and HygTk targeting vector compatible with the pEF-FS-EGFP RMCE system. The SEAP gene was PCR amplified with primers 36/ 37 on pSBC2SEAP (SB, HZI) and cut by BstBI/ MfeI. The vector pFS-

	EGFP-IRES-HygTk was also cut by BstBI/ MfeI, and dephosphorylated to exchange the EGFP gene by the SEAP gene.
pflpo-puro	Eukaryotic expression vector for the codon optimised recombinase FlpO and a puromycin resistance gene (Turan *et al.*, 2010).
pPGKMFGeGFP1	EGFP targeting vector of a RMCE system. The gene cassette is flanked by a wt and a F5 FRT site. It bears an EGFP gene between MLV 5' and MLV 3' sequences and a human PGK promoter at the 3' end followed by an ATG start codon. (Wirth, D., HZI, Braunschweig)
ptagSVHTGdneodC	Tagging vector of a RMCE system expressing HygTk. The SV40 promoter, the HygTk gene and a SV40 late polyA site are flanked by a wt and a F5 FRT site. There is an ATG-deleted neomycin phosphotransferase (Δneo^R) gene downstream of the F5 FRT site. (Wirth, D., HZI, Braunschweig)

2.2.10 Software

Text, graphic design and tables were generated by Microsoft software Word, Powerpoint and Excel and CoralDraw X3 (Adobe). Image editing was done by Photoshop 5.0 (Adobe). For the development of cloning strategies, PCR primer design, restriction- and sequence analysis Vector NTI suite 8.0 software (Invitrogen) was used. Pairwise and multiple sequence alignments were performed using CLUSTALW (Thompson, 1994) and EMBOSS (Rice *et al.*, 2000).

2.3 Molecular biology methods

Molecular-biological methods used in this work are adapted from standard collections of methods and protocols (Ausubel *et al.*, 2007; Coligan *et al.*, 2002; Sambrook & Russell, 2000).

2.3.1 Polymerase chain reaction (PCR)

Target genes from genomic DNA or template plasmids were amplified by polymerase chain reaction (PCR) to allow cloning into the respective recipient plasmid, analysis of cloned plasmids (Colony PCR) and analysis of eukaryotic high molecular weight (HMW) DNA. Standard PCR mixtures contained:

Table 2-7: PCR reagent mixes.

Component	Expand Hifi^{Plus} polymerase reagent mix	Taq polymerase reagent mix (Colony-PCR)
Primer forward (10 µM)	0.4 µM	0.3 µM
Primer reverse (10 µM)	0.4 µM	0.3 µM
Reaction buffer	20% (v/v) (1.5 mM $MgCl_2$)	10% (v/v) (2 mM $MgCl_2$)
Betaine (5M)	-	1.25 M
dNTPs (10 mM)	0.2 mM	65.5 µM
DMSO	5% (v/v)	-
Polymerase	2.5 U – 5 U	1 U
Template	10 pg – 10 ng (plasmid DNA) 50 ng – 100 ng (HMW DNA)	DNA from one *E. coli* colony
Water (MilliQ purified)	add up to 20 – 50 µl	add up to 20 µl

The reaction was performed in a thermo cycler T-Gradient (Biometra). Typical runs were carried out as described in the polymerase manufacturer's protocol or as follows.

Table 2-8: Thermo cycler program. The Annealing temperature depended on the primers' Tm.

Program	T (°C)	Time	Cycles (n)	dT/dn(s)	Heating rate
Initial denaturation	94	2 min	1		3°C/s
Denaturation	94	30 s			3°C/s
Annealing	Tm – 5 to 10	30 s	5-10		3°C/s
Elongation	72	1 min/kb			3°C/s
Denaturation	94	30 s			3°C/s
Annealing	Tm – 5 to 10	30 s	10-20	10	3°C/s
Elongation	72	1 min/kb			3°C/s
Final elongation	72	7 min	1		3°C/s

Amplificates were analysed by agarose gel electrophoresis (2.3.2). If necessary, they were purified by gel extraction (2.3.3).

2.3.2 Agarose gel electrophoresis

DNA samples were analysed by agarose gel electrophoresis using gels prepared from 0.8 % (w/v) agarose in TAE, brought to boil in a microwave. After cooling to ~60°C, the agarose solution was supplemented with ethidium bromide (10 mg/mL stock solution, 1:20.000) for DNA visualization and poured into an electrophoresis chamber. Plastic combs with teeth of different width were used to create sample pockets of desired volume.

DNA samples were mixed with 6x DNA loading buffer before injection into the sample pockets. Gels were run in TAE buffer with 5-7 V/cm for 50-60 min and documented under UV illumination at 254 nm.

TAE buffer: 40 mM Tris
20 mM Sodium acetate
1 mM EDTA
adjusted to pH 8.2 with acetic acid

6 x DNA loading buffer:

70% (w/v)	Sucrose
0.25% (w/v)	Bromphenol blue
0.1 M	EDTA

2.3.3 Extraction of DNA from agarose gels

Linearised vectors and PCR products were extracted from agarose gels using the “Gel extraction Kit” (Qiagen) or Plasmid Extract II Kit (Macherey-Nagel) according to their manuals. DNA was eluted with 20 µL TE-buffer or purified water (MilliQ).

TE-buffer:

Tris-Cl, pH 7.5	10 mM
EDTA	1 mM

2.3.4 Digestion of DNA with restriction endonucleases

Plasmid DNA was digested by distinct combinations of restriction endonucleases as required for cloning and control purposes. Buffers and reaction conditions were adapted to the enzyme combinations, according to the New England Biolabs technical reference catalogue.

2.3.5 Dephosphorylation of linearised plasmid DNA

Linearised plasmid DNA was 5'-dephosphorylated prior to ligation to prevent religation of the original vector without the target insert. 1-5 µg of linearised vector per planned ligation reaction were incubated with 10 U calf intestine phosphatase (CIP, NEB) at 37°C for one hour prior to agarose gel-purification.

2.3.6 Ligation of DNA fragments

Target genes (inserts) were ligated into linearised, 5'-dephosphorylated vectors using 30 ng linearised plasmid DNA and a 3-fold molar excess of insert. The mixture was incubated with 1 U of T4 DNA ligase (NEB) in the recommended ATP-containing ligation buffer (NEB) at RT for 4-5 h or at 22°C for 2 h followed by 16 h at 16°C.

2.3.7 Ligation independent cloning

Ligation independent cloning was performed using the In-Fusion^{TM} Dry-Down PCR Cloning kit according to its manual (Clontech). Briefly, a PCR reaction generated inserts bearing 15 bp extensions at both ends that are homologous to the ends of a linearised recipient vector. When the In-Fusion enzyme is incubated (15 min at 37°C followed by 15 min at 50°C) with the linearised vector (100 ng) and the PCR product (2-fold molar excess) in a 10 µl reaction mix, the enzyme's exonuclease activity excises nucleotides from the 3' ends of the molecules. The exposed overlapping ends are free to anneal (single-strand annealing, SSA), forming non-covalently joined molecules that undergo final repair within the target *E. coli* strain after transformation (2.3.9). The resulting product is an assembled vector and insert.

2.3.8 Preparation of electrocompetent cells

Electro-competent cells were prepared using *E. coli* TOP10 cells freshly plated on LB-agar. A single colony was picked, transferred to 100 mL of LB-medium, and incubated o.n. (37°C, 180 rpm). 5 mL of this starter culture were transferred to 500 mL LB-medium. Bacteria were grown to an $OD_{600\ nm}$ of 0.4, cooled rapidly and kept on ice for 30 min. Cells were pelletized by centrifugation (4°C, 10 min, 5500 g), washed twice with ice-cold, sterile 1 mM HEPES buffer, once with ice-cold, sterile 10 % (v/v) glycerol and resuspended in a final volume of 1 mL 10 % (v/v) glycerol (concentration 1-3 x 10^{10} cells/mL). 60 µL aliquots were transferred into pre-cooled 1.5 ml reaction tubes

(Eppendorf) and flash-frozen in liquid nitrogen. The electrocompetent cells were stored at -70°C.

2.3.9 Transformation of competent bacteria

Competent bacteria were transformed for plasmid amplification. Positive transformants were selected by plating the bacteria onto LB-agar plates containing antibiotics for which resistance was conferred by the respective plasmid.

Electroporation

1-2.5 µL of plasmid DNA, ligation- or In-Fusion™ cloning mixture were pipetted into a cooled, sterile 0.2 cm electroporation cuvette. A slowly thawed 60 µL aliquot of electrocompetent cells was added and the mixture incubated for 1 min on ice. The cuvette was dried and placed into a Gene Pulser™ (BioRad). An electric pulse at 2.5 kV, 200 Ω and 125 µF was applied to the cells, followed by a rapid addition of 1 mL of room-tempered LB-medium. After incubating at 37°C with gentle shaking for 1 h, the transformants were plated on selective LB-agar and incubated at 37°C o.n..

Heat shock transformation

5-10 µL of plasmid DNA or ligation mixture were added to 50 µL of chemically competent cells thawed on ice and incubated for 30 min on ice. A heat shock of 45 sec at 42°C was applied to the mixture, followed by a rapid cooling on ice for 2 min. 800 µL of SOC-medium were added to the cells followed by incubation for 1 h at 37°C with gentle shaking. The transformants were plated on selective LB-agar.

2.3.10 Plasmid preparation

Depending on the amount of plasmid DNA required, 5 (“Mini“) or 500 (“Maxi“) mL of LB-medium were inoculated with freshly transformed *E. coli* TOP10, XL-1 blue or DH5α cells. Bacteria were grown o.n. at 37°C and 180 rpm and the plasmid DNA was isolated with the QIAprep Spin Kit (Qiagen) or EndoFree Plasmid Maxi Kit (Qiagen) basing on the alkaline extraction method by Birnboim and Doly (Birnboim & Doly, 1979). The latter kit was used, if the DNA was used for transfecting mammalian cells. Yields were determined spectro-photometrically (2.3.12).

2.3.11 Genomic DNA preparation

Genomic DNA was isolated from CHO Lec3.2.8.1 cells in exponential growth phase with the AquaGenomic™ kit (MoBiTec). Yields were determined spectro-photometrically (2.3.12).

2.3.12 Determination of DNA concentration

Samples of isolated bacterial plasmid DNA were measured on a NanoDrop ND-1000 photospectrometer to determine the yield and purity. The absorption at 260 nm ($A_{260\ nm}$) was determined, allowing the DNA concentration to be calculated using Equation 2-1 and assuming that $A_{260\ nm}$ = 1.0 is equivalent to 50 µg/mL DNA.

$$c\ \left[\frac{\mu g}{\mu l}\right] = \frac{A_{260nm} \times D}{20} \qquad \textbf{Eq. 2-1}$$

$A_{260\ nm}$ Absorption at 260 nm
c concentration of DNA solution
D dilution factor

The ratio of $A_{260\ nm}$ /$A_{280\ nm}$ was used to check the sample for protein contamination. Ratios of 1.8 to 2.0 were assumed sufficiently pure to allow subsequent cloning and transformation/transfection experiments.

2.3.13 DNA sequencing

DNA sequencing reactions were performed by the company GATC or in the department of Genome Analysis at the HZI. The sequences were determined with the dye terminator method and provided as ABI chromatogram files for download. Chromatograms were analysed and compared using the programs AlignX and ContiqExpress from Vector NTI suite 8.0 (Invitrogen).

2.3.14 Southern Blotting

Genomic DNA was analysed by Southern Blotting in order to identify the copy number of the integrated transgene. 10 µg CHO DNA were digested with appropriate restriction enzymes and 0.25 µg/µL RNAse A. The digested DNA was separated in 0.8% agarose gels with a digoxigenin-labelled DNA molecular weight marker II (Roche) and transferred to Hybond N^+ membranes (Amersham) o.n. by capillary transfer with 20 x

SSC buffer. After washing the membrane in 2 x SSC buffer, the transferred DNA was fixed to the membrane by UV-cross-linking. For hybridisation and immunological detection the DIG High Prime DNA Labelling and Detection Starter Kit II (Roche) was used according to its instruction manual.

DIG-labelled DNA probes (Table 2-9) were generated by random prime labelling technique. 200-500 ng DNA probes in 16 µl purified water (MilliQ) were boiled 10 min at 99°C, cooled on ice and filled up to 20 µl with DIG High Prime mix (Roche) for o.n. incubation at 37°C. The reaction was stopped by adding 2 µl 0.2 M EDTA (pH 8.0) before the labelled probes were stored at 4°C.

Table 2-9: DIG-labled DNA probes.

Probe	Origin
EGFP probe	755 bp fragment (BglII/PstI) of pEpi-delCFGSARE
IRES probe	457 bp fragment (KpnI/PacI) of pFlpO-puro
PGK probe	667 bp fragment (NgoMIV/SacI) of pFS-scHGF-pGK

The membrane blots were pre-hybridized in DIG Easy Hyb solution (Detection Starter Kit II, Roche) for 1-2 h at 42°C. The denatured DIG-labelled DNA probe (5 min at 99°C, cooled on ice) was diluted in DIG Easy Hyb solution to 30 ng/mL and added without air bubble formation to the membrane. After o.n. incubation at 42°C, the membrane was washed under constant agitation - two times with 2 x SSC, 0.1% SDS buffer at RT and two times with pre-warmed 0.5 x SSC, 0.1% SDS buffer at 65°C for 15 – 30 min depending on the probe. For immunological detection the membrane was rinsed in water and unspecific binding sites were saturated for 30 min by Blocking Solution. The membrane was then incubated for 30 min in an anti-digoxigenin AP antibody solution (dilution 1:10,000 in Blocking Solution), washed three times for 15 min in Washing Buffer and equilibrated 5 min in Detection Buffer. To visualize the antibody on the blot about 3 ml CSPD ready-to-use substrate solution (Roche) were applied and the membrane was incubated 5 min at RT and 10 min at 37°C without drying. For documentation of the DNA bands, enhanced chemiluminescence (ECL) was detected by a luminometer.

20 x SSC buffer:	Tri-Natriumcitrate-2-hydrate	0.3 M
	NaCl	3 M

Blocking Solution:	10x Blocking Solution (Roche) diluted in Maleic acid NaCl adjusted with NaOH (solid) to pH 7.5	 0.1 M 0.15 M
CSPD:	0.25 mM Disodium 3-(4-methoxyspiro {l,2-dioxetane-3,2'-(5'-chloro)tricyclo[3.3.1.13,7]decan}-4-yl) phenyl phosphate	
Detection Buffer:	Tris-HCl/ pH 9.5 NaCl	0.1 M 0.1 M
DIG High Prime mix:	0.5 mM dATP, 0.5 mM dCTP, 0.5 mM dGTP, 0.325 mM dTTP, 0.125 mM DIG-11-dUTP, random hexanucleotides, DNA polymerase I (klenow enzyme large fragment) (Roche)	
Washing Buffer:	Maleic acid NaCl Tween 20 adjusted with NaOH (solid) to pH 7.5	0.1 M 0.15 M 0.03 % (v/v)

2.4 Cell culture

2.4.1 Maintaining cells in culture

Mammalian cells were propagated in suspension cultures of ProCho5 medium (Lonza, Cologne, Germany) supplemented with 7.5 mM L-glutamine and 11 mg/mL phenolred at 37°C in a humidified atmosphere of 5% CO_2. 125 mL- spinner flasks (Techne) were filled with a volume of 35 - 40 mL and 1000 mL-spinner flaks were filled with a volume of 200 - 250 mL and stirred at 80 rpm. 6-well plates were filled with a culture volume of 3 mL, T-25 cell culture flasks (25 cm^2, BD Falcon) were filled with 5 mL and T-75 cell culture flasks (75 cm^2) were filled with 20 ml. These vessels were incubated at 37°C in a humidified atmosphere of 5% CO_2 at 150 rpm on an Incutec K15-500 linear shaker. To maintain the cells in the exponential growth phase, they were subcultured in fresh medium every 3 to 4 days with a seeding cell density of 1.5-3 x 10^5 cells/mL. Cell numbers and viability were assayed by trypan blue dye exclusion method and/or Casy Counter (2.4.4).

Cells were cultured adherent in CD-Hybridoma medium supplemented with 5% FCS and 8 mM L-glutamine at 37°C in a humidified atmosphere of 8% CO_2. Cultures at about 80% confluence were split (1:5) or expanded as follows. After removal of the growth medium, cells were trypsinised 5 min at 37°C by 50 – 500 µL trypsin/EDTA solution (PAA laboratories) depending on the culture dish format. 0.2 – 2 mL medium

were added to the detached cells and the cells were pelletized at 180 g for 5 min and resuspended in a proper volume of medium.

2.4.2 Cryopreservation

For long-term storage mammalian cells were harvested and centrifuged at 180 g for 4 min. The cell pellet was resuspended in freezing medium (80% (v/v) culture medium, 10% DMSO and 10% Methylcellulose (1%)) at a cell density of 3 x 10^6 – 1 x 10^7 cells/mL. Each cryo-tube (Nunc) was filled with 1.8 mL cell suspension and stored in a special freezing container (Nalgene) in isopropanol at -70°C for about 24 h. Then, the cryo-tubes were transferred in the vapour phase over liquid nitrogen.

2.4.3 Revitalisation

For revitalisation the cells were thawed in a water bath at 37°C. Cells were washed two times with 10 mL pre-warmed culture medium (180 g, 4 min). The final pellet was resuspended in 3 to 40 mL culture medium depending on the cell numbers that had been frozen. Cell numbers and viability were assayed by trypan blue dye exclusion method (2.4.4) and cell density was adjusted to 2-5 x 10^5 cells/mL.

2.4.4 Determination of cell number and viability

The cell number and viability was determined by the trypan blue exclusion method or by a CASY particle counter.

For the trypan blue exclusion method cells were mixed 1:1 with 0.5% trypan blue solution (Roche). The cells were counted in four big squares of a Neubauer hemocytometer (BLB, Braunschweig) under a light microscope. The cell density (X) is calculated as follows:

$$X = \frac{C \times D}{4} \times 10^4 \qquad \textbf{Eq. 2-2}$$

X = cell density [cells/mL]

C = Cell numbers of four big squares

D = Dilution factor

While trypan blue can not enter the cytoplasmic membrane of viable cells, it stains selectively dead cells. The ratio of the viable cell number to the total cell number describes the viability of a culture:

$$Viability = \frac{c_v}{c_t} \times 100\ [\%]$$ **Eq. 2-3**

C_v = viable cell density [cells/mL]

C_t = total cell density [cells/mL]

The CASY particle counter is an electronic device based on resistance measurements. The cell suspension was diluted in an isotonic buffer and drawn in a capillary with defined diameter. When a cell passed two platinum electrodes, the resistance rose and resulted in electric pulses which corresponded to the cell number. In terms of pulse height a cell size profile was generated. The viability was assessed based on the integrity of plasma membrane. Injured cell membranes lose their function as electric insulators thus dead cells appear smaller than viable ones.

2.4.5 Quantification of glucose and lactate

The glucose and lactate concentration of culture supernatants was determined by a biochemical analyser (YSI) based on enzymatic-amperometric processes.

To quantify the glucose, 20-50 µl culture supernatant were pumped into a measuring cell. The cell was bordered by a probe containing glucose oxidase (GOD) immobilised on a membrane. Thus, the glucose reacts as follows:

I $\quad glucose + O_2 \xrightarrow{GOD} H_2O_2 + gluconic\ acid$

II $\quad H_2O_2 \xrightarrow{platimum\ electrode} 2H^+ + O_2 + 2\ e^-$

The resulting two electrons induce an electric current proportional to the glucose concentration in the sample.

For the determination of the lactate concentration the enzyme L-lactate-oxidase (LOD) was immobilised. The culture supernatant was pumped into the measuring cell and lactate was transformed as follows:

I $\quad lactate + O_2 \xrightarrow{LOD} H_2O_2 + pyruvate$

II $\quad H_2O_2 \xrightarrow{platimum\ electrode} 2H^+ + O_2 + 2\ e^-$

The resulting two electrons induce an electric current proportional to the lactate concentration in the sample.

2.4.6 Gene transfer methods

Nucleofection (Amaxa)

For stable transfection, 1.5 x 10^6 cells in 2 mL medium were transfected with *SalI* linearised plasmid DNA (5 µg) by using program U-24 of the Amaxa nucleofection device according to manufacturer's guidelines (Nucleofector™ Kit V, Lonza). 24 h post transfection the medium was exchanged and the cells were seeded into 6-well plates at 37°C in a humidified atmosphere with 5% CO_2 at 150 rpm on an Incutec K15-500 linear shaker. In the following days the transfected cell cultures were expanded before entering the stationary phase.

Recombinase-mediated GFP excision

Tagged cells were transfected with 5 µg of the optimized Flp expression vector pPGKFLPobpA by Amaxa nucleofection. 5 days later, the cells were analysed by flow cytometry and cloned by serial dilution in 96 well plates with CD-Hybridoma medium (2.4.7). After one week, the wells were screened for non-fluorescent single colonies on an Axiovert100 fluorescent microscope (Carl Zeiss) using an LP520 filter for GFP visualization. The detected colonies were expanded and recloned, if necessary. Clonal cell lines were adapted to suspension cultures in serum free ProCho5 medium by adding 10 U/ml heparin (Sigma) during the first 2 passages.

Recombinase-mediated cassette exchange

The cassette exchange in tagged CHO Lec3.2.8.1 cells was performed by Amaxa nucleofection. The cells were co-transfected with 1 µg, 2 µg or 2.5 µg of the targeting vector and 4 µg, 3 µg or 2.5 µg of the optimized Flp expression vector pPGKFLPobpA, respectively. 24 h post transfection the cells were seeded on a 100 mm culture dish in 10 mL CD-Hybridoma medium and cultivated at 37°C in a humidified atmosphere with 8% CO_2. Depending on the RMCE system 2 mg/mL G418 or 150 µg/mL Hygromycin B was added 5 days post transfection to select for the targeted subcell clones. Medium was replaced every 3-4 days until the subcell clones were picked in 96 well plates after 2-3 weeks. If necessary, the cells were recloned by serial dilution (2.4.7). Clonal cell lines were adapted to suspension cultures in serum free ProCho5 medium by adding 10 U/ml heparin (Sigma) during the first 2 passages.

2.4.7 Single cell cloning by serial dilution

Serial dilution in 96 well flat bottom microtiter plates was used to isolate clones of single cell origin. Initially, one well (A1) was inoculated with 0.5 x 10^4 cells/mL in 400 µL CD-Hybridoma medium. These cells were diluted 1:2 along the first column resulting in 200 µL per well. The second 1:2 dilution series was across the entire plate, so that all wells ended up with a volume of 100 µL medium and cells. The cells were cultured about one week at 37°C in a humidified atmosphere with 8% CO_2 until they were expanded or recloned again.

2.4.8 Flow cytometry and preparative FACS

CHO Lec3.2.8.1 cells were transported and sorted at room temperature. GFP expression was analysed with a Guava EasyCyte™ Mini System (488 nm laser, 520 nm bandpass filter, Guava Technologies). Cells were diluted to 2-4 x 10^5 cells/mL in PBS and stained with 50 µg/mL propidium iodide (PI) to exclude dead cells from analysis. Preparative FACS was performed on a MoFlo high-speed cell sorter (Beckman Coulter). The sorter was equipped with an argon-ion laser tuned to 488 nm with 100 mW of power and an automated cell deposition unit for sorting into 96-well plates. GFP fluorescence was detected in FL1 through a 530/40-nm bandpass filter. Data analysis was performed using CytoSoft™ 4.2 and WinMDI 2.9 software.

PBS-buffer:	NaCl	140 mM
	KCl	3 mM
	Na_2HPO_4/ pH 7.4	10 mM
	KH_2PO_4	2 mM

2.4.9 Cell cycle analysis

Cell cycle analysis by flow cytometric measurements on a LSR II cell analyser (BD) involved the determination of relative DNA contents. 2 x 10^6 cells were fixed o.n. in 3.5 mL ~70% Ethanol at -20°C. To prevent clustering the cells were resuspended in PBS before the addition of 2.5 mL absolute Ethanol. After pelletizing (180 g, 5 min) the cells were incubated in 500 µl PI staining solution for 40 min at 37°C. 3 mL PBS were added, the cell suspension was centrifuged and the pellet was resuspended in 500 µl PBS for analysis. The flow cytometer was equipped with a laser tuned to 488 nm. PI fluorescence was detected in FL1 through a 685/35-nm PI filter using linear amplification. Data analysis was performed using FlowJo v8.8.7 software (TreeStar).

DNA index was calculated by taking the ratio of geometric mean PI fluorescence of the fitted G0/G1 peak in the sample to that in the internal control.

PI staining solution:	Propidium Iodide	50 µg/mL
	RNase A	0.1 mg/mL
	Triton X-100	0.05 %
	in PBS buffer	

2.5 Protein production and purification

2.5.1 Recombinant protein production in stirred tank reactors (STRs)

Production cell lines were grown in batch or batch followed by perfusion mode in autoclavable stirred tank bioreactors with 2.5-L culture volume. The bioreactors were equipped with a double membrane stirrer for both bubble-free aeration via 8 m of hydrophobic polypropylene membrane tubing (Accurel® S6/2, Membrana, Wuppertal, Germany) and perfusion with internal cell retention via 8 m of hydrophilized, microporous membrane tubing of the same type (Lehmann *et al.*, 1985, Blasey *et al.*, 1990, Figure 2-1). Cells were propagated in ProCHO5 + ZKT-I [1:1] medium at 37°C, a stirring speed of 45 rpm, pH 7.4 and a dissolved oxygen (DO) concentration of 40% air saturation. During the production phase, at a cell density exceeding 10^7 cells/mL, temperature was reduced to 32°C. The perfusion rate was adjusted to the metabolic consumption of glucose avoiding a drop below 2.5 g/L. 2-mL samples were taken daily for routine in-process control. Harvested cell-free supernatant was concentrated by ultrafiltration followed by diafiltration against PBS8 using a Pellicon 2 tangential flow system equipped with two 10-kDa cut-off cartridges (Millipore).

PBS8-buffer:	NaCl	300 mM
	Na_2HPO_4/ pH 8	50 mM

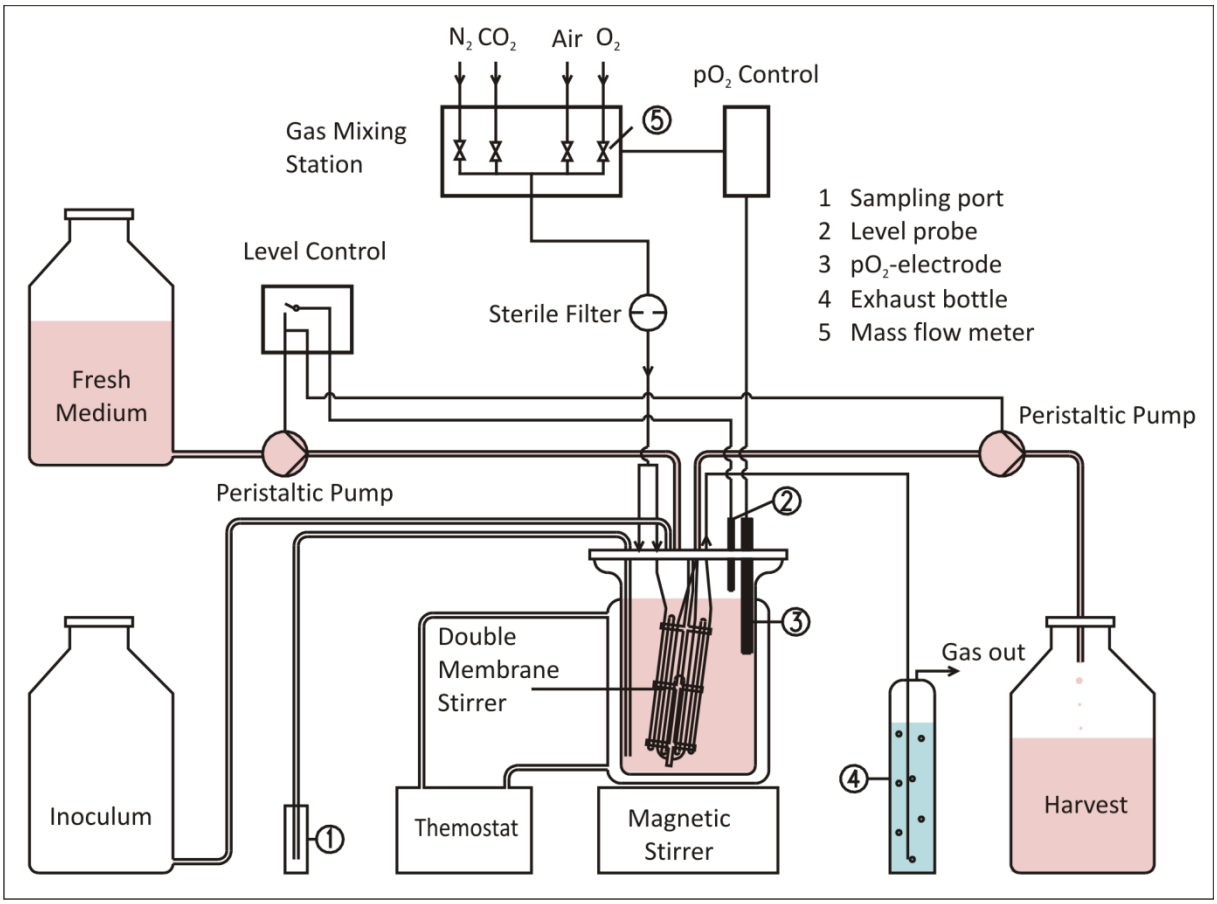


Figure 2-1: Flow diagram of a continuously perfused double membrane stirred tank reactor with bubble-free aeration for the production of recombinant glycoproteins by CHO cells.

2.5.2 Cell lysis

Intracellular proteins were extracted by the Cytobuster™ protein extraction reagent (Novagen). 1 x 10^6 pelletized cells (180 g, 4 min) were resuspended in 150 µL Cytobuster™ reagent. The cell suspension was incubated 45 min at room temperature and transferred to a 1.5 mL reaction tube for centrifugation (13,000 g, 5 min). The clear cell extract was transferred to a new reaction tube.

2.5.3 Affinity chromatography

The secreted human LAMP3-prox protein was fused to a hexa-histidine tag. Thus, the recombinant fusion protein was purified from the concentrated and diafiltrated culture supernatant (2.5.1) by immobilized metal ion affinity chromatography (IMAC) on a 20 mL HisPrep™ FF 16/10 (GE Healthcare) column using Äkta FPLC systems (GE Healthcare). All used buffers and H_2O were degassed and filtered (0.2 µm).

The column was equilibrated with 5 column volumes (CV) of binding buffer before the protein solution was applied by a flow rate of 0.9 mL/min. Immobilized proteins were

washed by binding buffer and flow rates not exceeding 0.3 MPa. Elution of the protein was achieved with a linear gradient from 0 – 300 mM imidazole within 12 CV finishing with a final step to 500 mM imidazole in 2 CV. 5 mL fractions were collected and OD_{280} was measured. Fractions absorbing stronger than buffer were analyzed by gel electrophoresis (2.6.2), pooled if appropriate and stored at 4°C.

Binding buffer:	NaCl	300 mM
	Na_2HPO_4/ pH 8	50 mM
Elution buffer:	Binding buffer + 500 mM Imidazole	

2.5.4 Gel permeation chromatography

Gel permeation chromatography (GPC) was used to separate molecules of different size in a porous gel matrix. GPC experiments were performed on ÄKTA FPLC Systems (GE Healthcare) with HiLoad Superdex 75 (formats 16/60, 120 mL and 26-60, 320 mL) columns in GPC buffer. All used buffers and H_2O were degassed and filtered (0.2 µm). The recombinant proteins were concentrated to 1-10 mL (2.5.6) and applied to the GPC column after equilibration with 1.2 CV H_2O and 1.2 CV GPC buffer. The protein separation was performed in 1 CV of GPC buffer with flow rates of 1 mL/min. Fractions of 1 – 5 mL were collected and OD_{280} was measured. Fractions absorbing stronger than buffer were analyzed by gel electrophoresis (2.6.2), pooled if appropriate and stored at 4°C.

GPC buffer:	HEPES/ pH 7.4	10 mM
	NaCl	150 mM

2.5.5 Deglycosylation of glycoproteins

Recombinant glycoproteins were deglycosylated at 1 mg/mL over night at 37°C by adding sodium acetate to 100 mM and endoglycosidase H_f to 10,000 U/mL (Endo H_f, NEB). Endo H_f is a fusion protein of endo H and maltose binding protein. Endo H_f and contaminants were removed by GPC (2.5.4).

2.5.6 Concentration of protein solutions

Protein solutions were adjusted to desired concentrations by ultracentrifugation using Vivaspin-2, -6, and -20 concentrators (Sartorius Stedim) with appropriate MW cut-offs. The pore size was chosen to be at least 10 kDa smaller than the target protein. The

absorption at 280 nm was checked against the flow-through until the desired protein concentration was reached (2.6.1).

2.6 Protein biochemical methods

2.6.1 Photometric quantification of protein concentration

Concentrations of purified protein solutions were determined via photometrical measurements of A_{280nm} against buffer. The molar extinction coefficient ε_{280} for each protein was calculated *in silico* with the program Vector NTI (Invitrogen) (Table 2-10). The concentration was calculated according to the Beer-Lambert law:

$$c = \frac{M_r}{\varepsilon d} A_{280} \qquad \textbf{Eq. 2-4}$$

c = concentration [mg mL^{-1}]
Mr = molecular mass [mg mmol-1]
A_{280} = absorption at λ = 280 nm
ε = molar extinction coefficient [M^{-1} cm^{-1}]
d = layer thickness of cuvette [cm]

Table 2-10: molar extinction coefficient ε.

Protein	ε_{280} [M^{-1} cm^{-1}]
hLAMP3-prox	9440
scHGF	148300

2.6.2 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Proteinaceous samples were analysed by SDS-PAGE with regard to their size and under denaturing conditions. The samples were mixed with 2 x Laemmli buffer and incubated for 5 min at 95°C to ensure complete unfolding. After a short centrifugation, 1 – 20 µl of the samples were applied to a SDS-polyacrylamide gel. The composition of these gels is shown in Table 2-11. For native gel electrophoresis the SDS component was omitted in the gel, in the running buffer and in the 2 x Laemmli buffer. In the 2 x Laemmli buffer β-Mercaptoethanol was omitted, too.

Table 2-11: Composition of SDS-polyacrylamide gels. Volumes are listed for four gels.

Solution	Separating gel		Stacking gel
	10%	12%	5%
Acrylamide/Bisacrylamide 30% (w/v)/0.8%(w/v)	10 mL	12 mL	1.5 mL
4 x Lower buffer	7.6 mL	7.6 mL	-
4 x Upper buffer	-	-	2.5 mL
10% SDS	300 µL	300 µL	-
H_2O	12.4 mL	10.1 mL	5.9 mL
TEMED	40 µL	40 µL	30 µL
40% (w/v) APS	60 µL	60 µL	30 µL

Gels were run at 120 V for 5 min and 160 V for 40-60 min, depending on the polyacrylamide percentage and the expected range of MW of the proteins present in the samples.

2x Laemmli buffer:	4x Upper buffer	1.25 mL
	10% SDS	1 mL
	Glycerol	2 mL
	β-Mercaptoethanol	200 µL
	Bromphenol blue (5mg/mL in 50% Glycerol)	500 µL
	H_2O	5 mL
4 x Upper buffer:	Tris-base adjusted with HCl, pH 6.8	0.5 M
	SDS	0.4 % (w/v)
4 x Lower buffer:	Tris-base adjusted with HCl, pH 8.8	1.5 M
10 x SDS-PAGE running buffer:	Tris-base	30.2 g
	Glycine	144.1 g
	SDS	10 g
	add with H_2O up to 1 L (pH 8.3)	

2.6.3 Coomassie staining

Unspecific staining of proteins in SDS-gels was performed o.n. in 20 - 30 mL staining solution. The gels were destained by two to three washing steps for about 20 min with destaining solution and H_2O before drying between cellophane films.

Staining solution:	Methanol	50 % (v/v)
	Acetic acid	10 % (v/v)
	Coomassie R-250	0.2 % (w/v)

Destaining solution:	Methanol	50 % (v/v)
	Acetic acid	10 % (v/v)

2.6.4 Western Blotting

For detection of specific proteins from culture supernatant Western Blots were performed followed by Immunostaining (2.6.5). Protein samples in SDS polyacrylamide gels were transferred to and immobilized on polyvinylidene difluoride (PVDF) membranes (Millipore) by a semi dry procedure. Freshly run gels were equilibrated for 15 min in transfer buffer, together with 2 pieces of gel-sized Whatman paper. The PVDF-membrane was activated in 100 % methanol for 1 min and briefly equilibrated in transfer buffer. For the transfer, the gel was placed onto the membrane, and both placed between two layers of soaked Whatman paper onto the anode of the blot apparatus. The blot was run with 15 V for 30 min. Success of the transfer was checked via reversible staining of the membrane with Ponceau red.

Transfer buffer:	Tris-base/ pH 8.0	25 mM
	Glycine	192 mM
	Methanol	15 %
Ponceau red staining solution:	Ponceau S	0.2 %
	Acetic acid	2 %

2.6.5 Immunostaining of Western Blots

Immobilized target proteins were detected using specific antibodies directed against His_6-tag of fusion proteins or directed against an epitope of human HGF/SF (Table 2-12). After the Western Blot (3.6.4) unspecific binding sites for antibodies were saturated by incubating the membrane for 1 h with 50 mL of 5% (w/v) skim milk in TBST at 37°C and washed 3 times for 5 min in TBST. The membrane was then incubated with 10 mL of the primary antibody diluted in TBST at 4°C o.n.. After antibody-binding, the membrane was washed 3 times for 5 min in TBST. Except for the membrane incubated with the peroxidase-conjugated antibody, a secondary antibody directed against the primary antibody and conjugated to an alkaline phosphatase (AP) was added. After 2 h incubation at RT, the membrane was washed 3 times for 5 min in TBST and 5 min in AP-buffer. To visualize the phosphatase-bound antibody 10 mL of a BCIP-NBT-solution were added (25 mg/mL BCIP, 50 mg/mL NBT in AP-Puffer; Applichem). After up to 30 min the reaction was terminated by a further washing step in water.

The peroxidase-bound antibody was visualized on the membrane using Lumi-Light Western Blotting Substrate (Roche). For documentation, protein bands were detected by enhanced chemiluminescence (ECL) on a luminometer.

Table 2-12: Primary and secondary antibodies for immunostaining of Western blots.

Primary antibody	Company	D	Secondary antibody	Company	D
mouse α-His_6-Peroxidase	Roche	1:1000	-	-	-
mouse α-His	Novagen	1:1000	rabbit anti-mouse AP	Promega	1:2000
goat α-HGF	R&D	1:1000	goat anti-mouse AP	Promega	1:1000

TBST-buffer:	Tris-base/ pH 8.0	20 mM
	NaCl	150 mM
	TWEEN 20	0.05 % (v/v)
AP-buffer:	Tris/ pH 9.5	100 mM
	NaCl	100 mM
	$MgCl_2$	5 mM

2.6.6 Enzyme-linked immunosorbent assay (ELISA)

scHGF product concentrations from cell culture supernatants were quantified with the human HGF DuoSet® ELISA Development System (R&D Systems) according to the manufacturers' protocol. Cells were seeded at 1.5 x 10^5 cells/mL in 3 mL in 6-well plates or 40 mL in spinner flasks. scHGF concentration was measured twice, one and four days after inoculation. Mean values were obtained from three parallel cultures. Daily cell culture supernatants from bioprocesses producing scHGF were analysed as a routinely in-process control, too.

For the analysis 96-well immuno-plates (F96 MaxiSorp™ surface, Nunc) were coated by a capture antibody (mouse α-human HGF, 1 µg/mL in reagent diluent) o.n. at RT. The wells were washed three times with wash buffer before blotting the inverted plate against paper towels. The unsaturated binding sites were blocked for 1 h at RT by 300 µL reagent diluents per well. The washing steps were repeated and 200 µL of the samples (1:2 diluted in PBS) and the HGF standard (42 ng/mL in PBS) were loaded in duplicate in the wells of the first line of the plate. Each sample and the standard were diluted 1:2 in PBS along the plate ending with 100 µL proteinaceous solution per well. After incubating 2 h, the wells were washed three times again and 100 µL detection antibody (biotinylated goat anti-human HGF, 200 ng/mL in reagent diluent) was loaded per well. The plates were incubated for 2 h, washed three times as before and incubated 20 min with 100 µL Streptavidin-HRP per well. Threefold washing removed

the excess of Streptavidin-HRP. For colour reaction 100 µL TMB plus ready-to-use substrate (KemEnTec Diagnostics) were added per well. After 2 min the reaction was stopped by 50 µL of 2N H_2SO_4. The plates were analysed at wavelengths of 450 nm and 540 nm (for correction) by a microplate ELISA reader (Infinite M200, Tecan). The scHGF concentration in the culture supernatant was calculated by the standard curve of HGF.

After counting the cell number, the specific productivity q_p [pg per cell per day; pcd] was calculated with the assumption of exponential growth by Equation 2-5 (Pirt, 1975).

$$p(t) = p(0) + \frac{q_p x_0 (e^{\mu t} - 1)}{\mu} \qquad \textbf{Eq. 2-5}$$

p = product concentration at time t since inoculation [µg/mL]

x_0 = inoculum cell concentration [cells/mL]

µ = specific growth rate [h^{-1}]

PBS:	NaCl	137 mM
	KCl	2.7 mM
	Na_2HPO_4/ pH 7.4	8.1 mM
Wash buffer:	PBS	
	+ Tween 20	0.05 %
Reagent Diluent:	PBS	
	+ BSA	1 %

All buffers were filtered (0.2 µm).

2.6.7 Fluorescence spectrometry

The EGFP protein concentrations in CHO Lec3.2.8.1 cell extracts were quantified by fluorescence spectrometry with an Infinite M-1000 fluorescence spectrometer (Tecan). Cells were seeded at 1.5 x 10^5 cells/mL in 3 mL in 6-well plates or 40 mL in spinner flasks. Fluorescence from the cell extracts (3.5.2) was measured in duplicate (excitation 470 nm, emission 510 nm) in non-fluorescent microtiter plates (Nunc) using several dilutions of the samples, and compared to a recombinant EGFP (BioVision). Linear regression of the values obtained for defined concentrations of recombinant EGFP (1 µg/mL, 0.8 µg/mL, 0.6 µg/mL, 0.5 µg/mL, 0.4 µg/mL, 0.2 µg/mL and 0.1 µg/mL) enabled the quantification of fluorescing protein in the cell extracts.

2.6.8 Mass spectrometry

Protein samples were analysed by mass spectrometry (MS), by Dr. Manfred Nimtz and Undine Felgenträger (HZI, Braunschweig).

Recombinantly produced, purified and eventually deglycosylated proteins were investigated by MALDI-TOF (Matrix Assisted Laser Desorption Ionisation – Time Of Flight) on an Ultraflex mass spectrometer (Bruker) or ESI-MS/(MS) (Electrospray Ionisation) on a QTOF II mass spectrometer (Micromass) regarding their glycosylation and/or intactness of the expected amino- and carboxy-terminal ends.

Protein preparation for MS analysis

To prepare desalted samples of native scHGF, 50 µL Ni-NTA superflow beads (Qiagen) were added to samples of His_6-tag scHGF. scHGF was bound by shaking for 1 h and unbound material was removed upon centrifugation. Beads were washed three times in 5 mM Tris-HCl, pH 8.0 and scHGF was eluted three times with 100-µL aliquots of 35% acetonitrile, 0.1% TFA and shaking for 10 min. Eluates were subjected to MALDI-TOF-MS analysis.

Protein crystals were desalted by three washing steps in 4 µL purified water (MilliQ). The crystals were dissolved by adding 2 µL methanol to the crystal in the water drop followed by mixing. The protein solution was subjected to MALDI-TOF MS analysis.

Protein bands from Coomassie-stained gels were excised, five times washed in purified H_2O (Milli Q) and tryptically digested by the group of Dr. Manfred Nimtz. Extracted polypeptides were desalted on reversed-phased C_{18} ZipTip pipette tips (Millipore Corporation) and subjected to ESI-MS/(MS).

2.6.9 N-terminal sequencing

Protein samples were identified and evaluated by N-terminal sequencing. All sequencing experiments were carried out by Rita Getzlaff (HZI, Braunschweig) as described (Edman & Begg, 1967).

2.6.10 Dynamic light scattering (DLS)

Dynamic light scattering (DLS) is used to measure hydrodynamic radii of proteins in solution based on their diffusion velocity. Particles scatter light in all directions resulting in intensity changes over time depending on the Brownian movement of the scattering

particles. These changes in light scattering intensity over time are evaluated with an autocorrelation function which allows calculation of the hydrodynamic radii of the analyzed proteins.

Besides hydrodynamic radii, polydispersity was calculated. Monodisperse and homogenous protein samples for crystallization typically have a polydispersity <15 %. Protein solutions were centrifuged (20 min at 13,000 rpm), diluted with GPC buffer up to 2 mg/mL and filtered with a 0.2 µm filter. All DLS measurements were performed at RT with a DynaPro Titan (Wyatt Technologies).

2.7 Protein crystallization

For X-ray diffraction experiments proteins had to be crystallized to obtain a regular arrangement of protein molecules in the three dimensional space. A crystal is set up by a regular arrangement of the unit cell, which is the smallest element to form a crystal based on translation along a crystal lattice. As the unit cell may contain symmetry elements itself, it can be divided into asymmetric units. The asymmetric units can be transformed into each other by the symmetry operations implied by the unit cell. The number of protein molecules per asymmetric unit may differ between different crystal forms and proteins. Unit cells are generally assigned to seven different crystal systems: triclinic, monoclinic, hexagonal, cubic, rhombic, tetragonal and orthorhombic. The packing is divided into one primitive and a number of centered ones (*C*: *c*-face centered; *I*: body-centered; *F*: all-face centered), such that combining both, the crystal systems and the packing, leads to 14 different Bravais lattices.

There are 230 different crystallographic space groups. However, only 65 account for protein crystals as any space group that contains symmetry elements (e.g. mirror planes, inversion centre, etc.) which change the chirality of the molecules is not allowed for proteins.

2.7.1 Initial screening for lead crystallization conditions

LAMP-3 protein solution was initially screened for lead crystallization conditions using commercial screens (Table 2-13) for 96-well sitting-drop vapour diffusion method. 200 nL drops composed of equal volumes of 10 mg/mL protein in GPC buffer and crystallization solution were pipetted using a MOSQUITO robot (TTP LabTech Ltd.). The plates were sealed with Manco^{TM} Crystal Clear tape (Jena Bioscience) and incubated at 19°C.

Table 2-13: Crystallization Screens.

Screen	Company
JCSG core I	Qiagen
JCSG core II	Qiagen
JCSG core III	Qiagen
JCSG core IV	Qiagen
The Classics	Qiagen
The Pegs	Qiagen

2.7.2 Optimization of lead crystallization conditions

Initial hit conditions were reproduced and optimized manually in 24-well hanging drop vapour-diffusion formats using 1-2 µL crystallization drops and 500 µL reservoir. Suitable crystals for X-ray diffraction experiments were obtained by varying physico-chemical parameters of the initial condition 0.1 M citric acid pH 5, 5% PEG 6000. Different precipitant concentrations between 3 – 14 % PEG 6000, protein concentrations of 7.9 mg/mL, 10 mg/mL and 22 mg/mL, pH values between 4 – 6 and incubation at 4°C, 19°C and 26°C were tested. Additionally, co-crystallization with additives (additive screen, Hampton Research Corp.) was applied.

2.7.3 Seeding

Seeding is used to improve crystal quality by introducing small nucleation seeds into crystallization droplets, setting a starting point for crystal growth. This can be obtained either by streak seeding, or by micro-seeding.

Streak seeding

A horse hair was pulled through a crystallization droplet containing crystals of hLAMP3-prox. Small crystal fragments stuck to the horse hair and were transferred to crystal free crystallization droplets by streaking through.

Micro-seeding

Some hLAMP3-prox crystals in 5 µL mother liquor were transferred to 45 µL reservoir and mixed thoroughly until the crystals were broken into small fragments. Serial dilutions from 10^{-1} up to 10^{-4} were set up from this stock solution. Crystallization droplets were then set up using 1 µL of the diluted micro-seeding solutions and 1 µL of purified protein (7.9 mg/mL in GPC buffer) and equilibrated against 500 µL reservoir in a 24 well crystallization plate (Greiner).

2.7.4 Heavy atom derivatisation

Heavy-atom derivatized LAMP-3 crystals were needed to solve the phase problem mentioned in 2.8. To find appropriate heavy atoms interacting with hLAMP3-prox a band shift assay was performed (Boggon & Shapiro, 2000). To this end, 1 µL of hLAMP3-prox (7 mg/ml) was incubated in 1 mM heavy atom solution in 0.1 M citric acid pH 4.5 and a final reaction volume of 10 µL for at least one hour. The reaction solution was analysed on a 15% native gel by electrophoresis (2.6.2). The following heavy atom compounds were screened: Kt_2PtCl_4, $Hg(OOCCH_3)_2$, $(C_2H_5HgO)HPO_2$, $(NH_4)_3IrCl_6*xH_2O$, $GdCl_3*xH_2O$, $Sm(O_2C_2H_3)_3*xH_2O$ and $AuCl_3$.

The crystals were soaked in mother liquor supplemented with millimolar heavy-atom compound concentrations according to Table 2-14. Crystals treated with heavy atoms were flash-frozen in liquid nitrogen.

Table 2-14: Heavy-atom compounds for derivatisation of hLAMP3-prox crystals.

Heavy-atom compound	Concentration in mother liquor	Soaking time
Potassium tetrachloroplatinate(II) (K_2PtCl_4)	25 mM	5 min
Ammonium hexachloriridate(III)hydrate ($(NH_4)_3IrCl_6*xH_2O$)	14 mM	0.5 – 10 min

2.8 X-ray diffraction analysis

X-ray crystal structure analysis was used to determine the electron density distribution within the unit cell from diffraction images. X-ray diffraction experiments are based on Bragg's law describing total reflection on a set of parallel planes with the Miller indices *hkl* (Bragg *et al.*, 1913). The Miller indices *hkl* describe a set of parallel planes where *hkl* represent the number of parts into which the set of planes cut the a, b and *c* edges of the crystallographic unit cell. X-rays of wavelength λ are reflected at the angle θ only, if θ meets equation 2-6.

$$2d_{hkl} \sin\theta = n\lambda \qquad \textbf{Eq. 2-6}$$

d_{hkl} = interplanar spacing

θ = reflection angle

λ = wavelength

n = integer

X-rays are diffracted at the crystal lattice and Bragg's law is valid for all points of the reciprocal lattice on a sphere with the radius 1/λ around the origin of the real lattice, the so called Ewald sphere. Rotating the crystal during an X-ray diffraction experiment maximizes the number of points hitting the Ewald sphere causing more detectable reflexes on the X-ray detector. To result in a complete data set, crystallographic data have to be collected over an angular range that depends on the crystal's orientation and inherent symmetry.

X-rays scattered at the atoms of a crystal are characterised by the three parameters of an electromagnetic wave: frequency, amplitude and phase. Both, frequency and amplitude of an emerging x-ray can be determined from the according Bragg peak's position and intensity. The phase, however, is not detected and not available for calculation of the electron density distribution ρ (*xyz*). The Fourier transform of the crystal's electron distribution function ρ (*xyz*) is its structure factor F (*hkl*) and vice versa. The structure factor can be expressed as a function of *hkl*:

$$F\ (hkl) = \iiint_0^1 \rho\ (xyz) e^{2\pi i\ (hx+ky+lz)} dx\ dy\ dz \qquad \textbf{Eq. 2-7}$$

Since Bragg diffraction occurs in discrete directions, the integration of the electron density calculation can be replaced by a summation of individual structure factors:

$$\rho\ (xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)|\ e^{-2\pi i(hx+ky+lz-\alpha_{hkl})} \qquad \textbf{Eq. 2-8}$$

The lack of phase information makes the direct determination of the structure factors and hence the electron density impossible. The structure factor amplitude, however, is proportional to their intensity (Eq. 2-9) and can be determined.

$$I_{hkl} \sim |F_{hkl}|^2 \qquad \textbf{Eq. 2-9}$$

In this study, the phase problem was solved by anomalous dispersion of heavy-atom derivatives of hLAMP3-prox crystals. Anomalous dispersion occurs when the energy of the incident X-ray photon is close to the absorption edge/transition energy required to promote an electron to an unoccupied higher orbital or to eject it completely from an atom. The absorption of incident energy by the crystal results in a reduced intensity of the coherent scattering and a slight change in the phase of the scattered X-ray beams. This is expressed by the atomic scattering factor f of the heavy atom comprising now the normal scattering component ^{0}f and a non-negligible wavelength-dependent real (dispersive, f') and an imaginary (absorption, f'') correction term for the anomalous

scattering. The experimental determination of the actual shape of the absorption curve *via* fluorescence energy scan of the heavy atom-labelled crystal provides the precise values for f'' and of its derivative f'. As the absorption term f'' is always 90° advance in phase, the resonant/anomalous scattering causes a slight deviation from Friedel's Law (Friedel, 1913). Friedel pairs F_{hkl} and F_{-h-k-l} (short F^+ and F^-) are Bragg reflections related by inversion through the origin exhibiting the same magnitude of the structure factor intensity but opposite phase angles at normal scattering using wavelengths far from the absorption edge. In anomalous scattering, Friedel mates no longer have equal structure factor amplitudes (Bijvoet, 1954): $|F_{hkl}| \neq |F_{-h-k-l}|$ and their phase angles are no longer complementary. Near the absorption edge energy ("peak" wavelength), the imaginary anomalous component f'', which is proportional to the atomic absorption coefficient of the heavy atom, shows its largest value. The dispersive part f' of the anomalous signal shows its minimum at the inflection point of the scan. The structure factors of the derivative at these wavelengths differ significantly in their amplitudes compared to the native data (or the equivalent data of a derivative collected at a remote energy). It has been shown that even a single wavelength is sufficient to obtain good phase estimates and to solve novel structures (single-wavelength anomalous dispersion (SAD) (Dauter *et al.*, 2002). In SAD, the phase ambiguity is broken by density modification protocols. The correct enantiomorph (handedness) of the anomalous scatterer substructure is thereby revealed by comparison of the electron-density maps treated by density modification. The incorrect enantiomorph contains no image of the macromolecule, whereas the correct map features clear solvent boundary and a macromolecular-like density histogram. The positions of the anomalously scattering heavy atoms in the unit cell can be deduced from an anomalous difference Patterson map (Patterson, 1935), which is calculated using only the coefficients ΔF^2 derived from the anomalous structure factor amplitude differences of the Friedel pairs (and their symmetry mates) $\Delta F = |F^+| - |F^-|$, also called Bijvoet amplitude differences. Significant non-origin peaks in the three dimensional Patterson map correspond to interatomic vectors between anomalously scattering atoms only that are related by crystallographic symmetry and are found in two-dimensional slices of the map (called Harker sections, (Harker, 1936)). From the Harker peaks (u,v,w), the coordinates of the heavy atoms (x,y,z) in the unit cell can be calculated. The normal scattering structure factor amplitude and phase, together with f'' and its derivative f', allows the estimation of the phases of the anomalous structure factors F_A for the anomalous scattering substructure. This

allows a first approximation of the phase angles for the native protein structure factors F_T.

2.8.1 Cryoprotection of crystals

Protein crystals are sensitive to X-ray radiation leading to severe damage and thus resulting in data loss, especially when using strong synchrotron X-ray radiation for data collection. To reduce this damage, X-ray diffraction data are collected under cryogenic conditions in a stream of liquid nitrogen at 100 K. Ice formation was prevented by quick-soaks for 10 to 60 sec of the hLAMP3-prox crystals in cryo-protection solution containing mother liquor supplemented with 30% (v/v) PEG6000 or dehydrated oil (centrifuged in a SpeedVac o.n.). Crystals were stored in cap tubes filled with liquid nitrogen or mounted onto the goniometer using a pre-cooled cryo-tong and maintained at 100 K throughout the data collection using a nitrogen stream.

2.8.2 Data collection

X-ray diffraction data were collected on a rotating copper anode generator and an R-AXIS IV++ image plate detector (Rigaku) and at synchrotron beamline X12 (EMBL, DESY, Hamburg) on charge-coupled device detectors (MarCCD, Marresearch). Data collection strategies were calculated by MOSFLM (Leslie, 1992). For single-wavelength and multiple-wavelength anomalous dispersion (SAD and MAD) experiments data sets of hLAMP3-prox derivatives were collected peak, inflection and high remote energies, respectively, of corresponding heavy-atom absorption L(III)-edges. The wavelengths were chosen from fluorescence scan data and are listed in Table 2-15.

Table 2-15: Data collection at different wavelengths.

hLAMP3-prox derivatives	Wavelength [Å]	Light Source
Sulphur	1.771 (high remote)	Synchrotron beamline X12 (EMBL, DESY, Hamburg)
Platinum	1.067870 (high remote) 1.070450 (peak) 1.070730 (inflection)	Synchrotron beamline X12 (EMBL, DESY, Hamburg)
Iridium	1.100090 (high remote) 1.103710 (peak) 1.104200 (inflection)	Synchrotron beamline X12 (EMBL, DESY, Hamburg)

2.8.3 Data processing

hLAMP3-prox diffraction data were processed anomalously by MOSFLM (Leslie, 1992) or XDS (Kabsch, 2010b) determining the crystal parameters, assigning the space group, indexing and integrating the images' reflections. Twinning was analysed by the program SFCHECK (Vaguine *et al.*, 1999). The intensities were scaled and partial reflections merged with XSCALE (Kabsch, 2010a). Finally the data were converted to a MTZ file by XDSCONV (Kabsch, 2010a).

2.8.4 Phasing

Initial phasing of 3-wavelength MAD experiments were performed by Dr. Jörn Krauße using the AutoSol program within the PHENIX suite software package (Adams *et al.*, 2010; Terwilliger *et al.*, 2009) and Coot (Emsley & Cowtan, 2004). Figures were generated with Pymol (http://pymol.org).

3 Results

The present work describes the development of different site-specific recombination strategies for cell line development, leading to a decisive improvement in the production of glycoproteins for structural studies. CHO Lec3.2.8.1 cells, which have a positive track record in X-ray crystallography, were chosen as expression hosts. Site-specific recombination systems facilitated the exploration of pre-characterized, favourable chromosomal loci for transgene expression in these cells. First, genetic loci were randomly 'tagged' with a reporter gene and loci with a positive effect on reporter gene expression were identified. Depending on the strategy, the reporter gene was either excised (Flp-mediated excision, FLEx), activating a gene of interest located downstream, or exchanged against genes of interest by recombinase-mediated cassette exchange (RMCE). The chromosomal integration locus was not changed during the recombination step, except for the reporter gene coding sequence. Therefore, the expression ability of the resulting production subcell clones was highly predictable. This in turn essentially reduced the effort in screening for clones with appropriate expression level and genetic stability. Consequently, the time for developing production cell lines was greatly reduced.

Site-specific recombination was applied for the first time for glycoprotein production in structural biology. Production, characterization and crystallization of various target proteins confirmed that X-ray structure analysis of glycoproteins benefits from site-specific recombination with CHO Lec3.2.8.1 cells.

3.1 Generation of master cell lines

3.1.1 Tagging vectors

Cell line development described in this work relies on the Flp/FRT recombination system from *Saccaromyces cerevisiae*. Chromosomal loci in the host cell genome were tagged with FRT sites. These sites were integrated in three different vectors flanking a GFP reporter gene (Fig. 3-1, 2.2.8). The tagging vector $pEF_{F3}EGFP_{F3}MCS$ (Kaufman *et al.*, 2008) (Gen-Bank GU983383) bears two identical FRT sites (F3) and thus confers FLEx. This resulted in a switch to expression of a GOI cloned into the downstream multi cloning site (MCS). Coding sequences for the single chain mutant of HGF/SF (scHGF) and the membrane-proximal domain of human LAMP-3/DC-LAMP (hLAMP3-prox) were cloned into this vector. The tagging vectors pEF-FS-EGFP and pEF-FS-EGFP-dneo

were designed for RMCE. Two heterospecific FRT sites (F3 and F) that can not recombine with each other flank the GFP gene. In the presence of a second, identically flanked DNA sequence, cassette exchange can be mediated by Flp. The vector pEF-FS-EGFP-dneo additionally contains a selection trap, an ATG-deficient aminoglycoside phosphotransferase gene (Δneo^R) downstream of the FRT cassette. The Δneo^R gene is complemented and activated by correct cassette exchange events in order to confer resistance to geneticin (G418). Further, a 'promoter trap' is realized by the position of the EF-1α promoter upstream of the FRT cassette. An incoming promoter-less transgene can only be expressed upon correct cassette exchange.

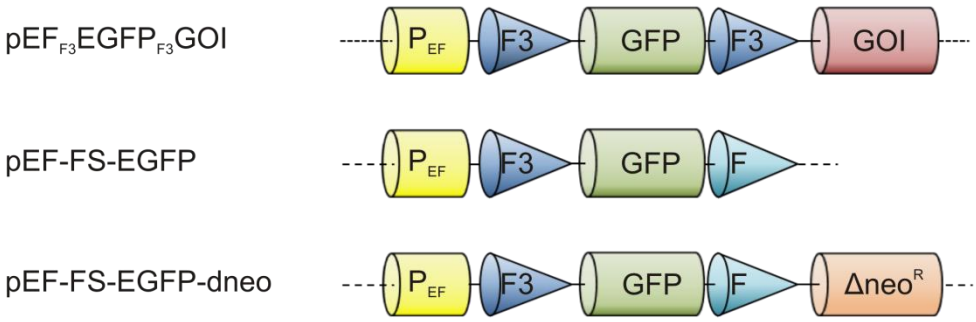


Figure 3-1: Schematic representation of tagging vector gene cassettes. pEF$_{F3}$EGFP$_{F3}$GOI is the tagging vector for the FLEx strategy; pEF-FS-EGFP and pEF-FS-EGFP-dneo are tagging vectors for the RMCE strategy. They differ by the additional ATG-deficient neomycin resistance gene (Δneo^R) downstream of the gene cassette on pEF-FS-EGFP-dneo. The Δneo^R gene is part of a selection trap. F = wild type FRT site; F3 = F3 mutant FRT site; P_{EF} = human EF1-α promoter.

3.1.2 Screening for expression loci in host cell genome

The general strategy to establish master cell clones for site-specific recombination is outlined in Figure 3-2. CHO Lec3.2.8.1 cells were transfected with the Nucleofector™ system ('nucleofected') with usually 5 µg linearised tagging vector. This created a broad spectrum of randomly tagged cells. Clonal 'master' cell lines that integrated the transgene into favourable chromosomal loci were established by selecting GFP positive cells by two rounds of preparative FACS. The isolated cells contained genetically stable chromosomal loci, tagged with the reporter gene and conferring high transcriptional activity. This strategy avoided the use of antibiotics for selection which was reported to improve the establishment of cell clones of homogenous and stable transgene expression (Kaufman *et al.*, 2008).

In the first sorting step, 7 days post nucleofection, GFP positive cells of the highest fluorescent were selected, corresponding to 1% to 50% of the cell population. Sorting was repeated two weeks after nucleofection and 3% to 11% of the cells were isolated

as single cells. Upon expansion, individual GFP master cell clones were obtained 28 days after nucleofection.

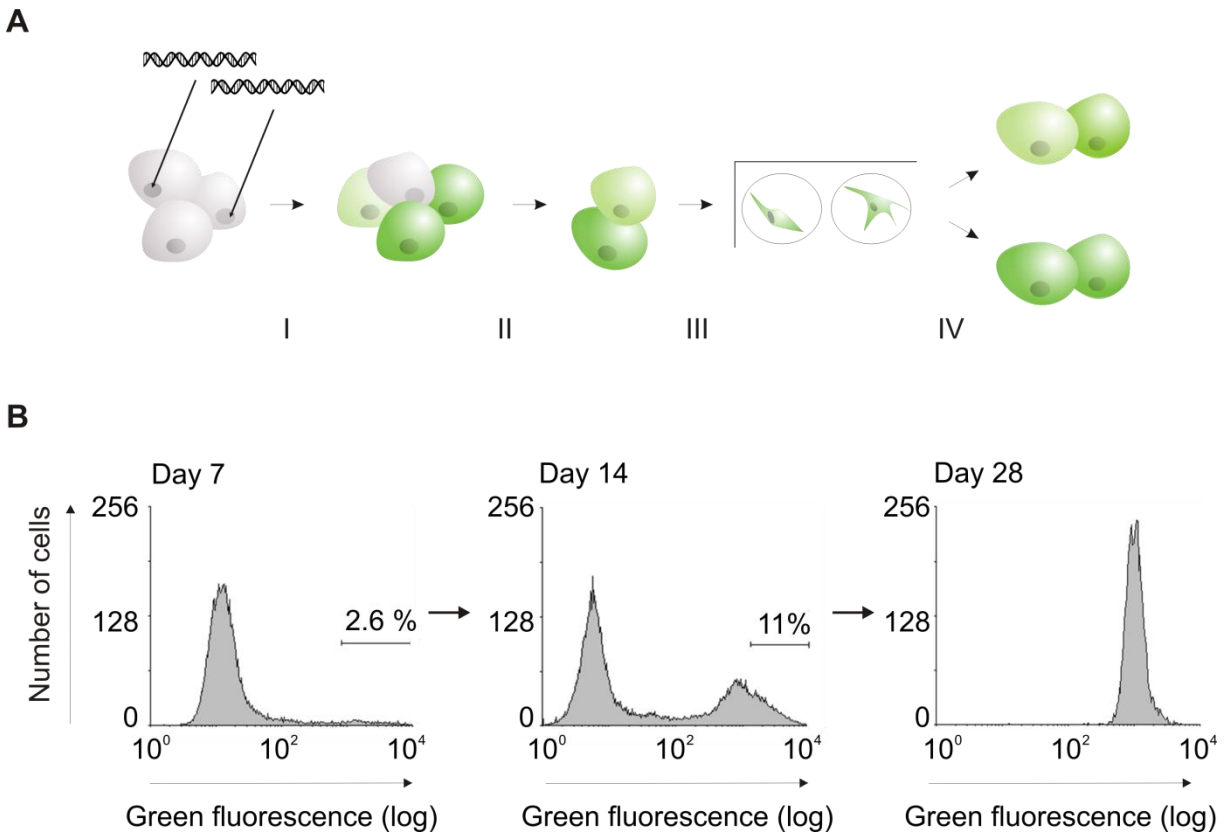


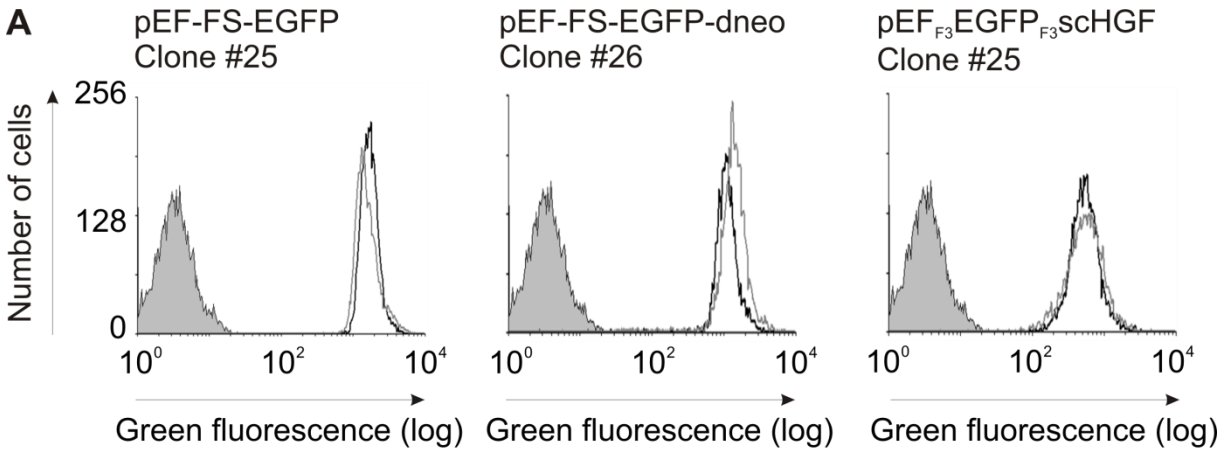
Figure 3-2: Generation of master cells using a FACS selection strategy. (A) Schematic representation of selecting tagged master cells for subsequent recombination. I: transfection of host cells with a GFP tagging vector. II: First sorting of top fluorescent GFP cells in a population by FACS one week post transfection. III: Second sorting of top fluorescent single cells out of the sorted population two weeks post transfection. IV: Expansion of single cell clones followed by the analysis of GFP fluorescence level. (B) Representative presentation of FACS profiles of transfected and sorted CHO Lec3.2.8.1 cells. Cells were transfected by pEF-FS-EGFP-dneo. Seven days post transfection the top 2.6% fluorescent cells were sorted in a population. On day 14 the top 11% fluorescent cells were sorted as single cells out of the sorted population. 28 days post transfection the cells had been expanded and the fluorescence profile of a representative tagged cell clone is shown.

3.1.3 Long term stability of tagged cell clones

An essential criterion for a production cell line is long term stability of transgene expression. Thus, potential master cells were cultured for 12 weeks and GFP fluorescence was analysed weekly by flow cytometry. Cell clones with stable GFP expression and favourable growth characteristics were eventually isolated for all tagging vectors (Fig. 3-3). Overlapping fluorescence profiles indicated stable expression. Cells with unstable GFP expression were usually identified within the first three weeks.

Unstable cell lines either segregated into GFP positive and negative populations, or GFP levels gradually shifted to lower fluorescence over time.

A



B

Tagging vector	Number of isolated clones	Number of stable clones
pEF$_{F3}$EGFP$_{F3}$scHGF	34	12
pEF$_{F3}$EGFP$_{F3}$LAMP3	46	18
pEF-FS-EGFP	54	30
pEF-FS-EGFP-dneo	60	15

Figure 3-3: Long term stability of master cells over 12 weeks. GFP flow cytometer profiles of representative master cell clones tagged by the vectors pEF-FS-EGFP (SWI2-25), pEF-FS-EGFP-dneo (SWI3-26) and pEF$_{F3}$EGFP$_{F3}$scHGF (SWI4-25), respectively. Cell clones were analysed when they were two weeks in culture (grey line) and when they were 12 weeks in culture (black line). CHO Lec3.2.8.1 cells served as negative control (grey plane). (B) Statistics of cell clones regarding their GFP expression stability.

To quantitate production strength, GFP concentrations were measured in triplicate from cell lysates of four-day cultures by fluorescence spectrometry. GFP concentrations ranged from 11 to 30 mg GFP per litre of culture. The concentrations were normalized to the viable cell number for a better comparison of unequally grown cultures (Fig. 3-4). Assuming that these cells can produce other proteins in a similar scale, they were considered as adequate producers for glycoproteins in structural biology.

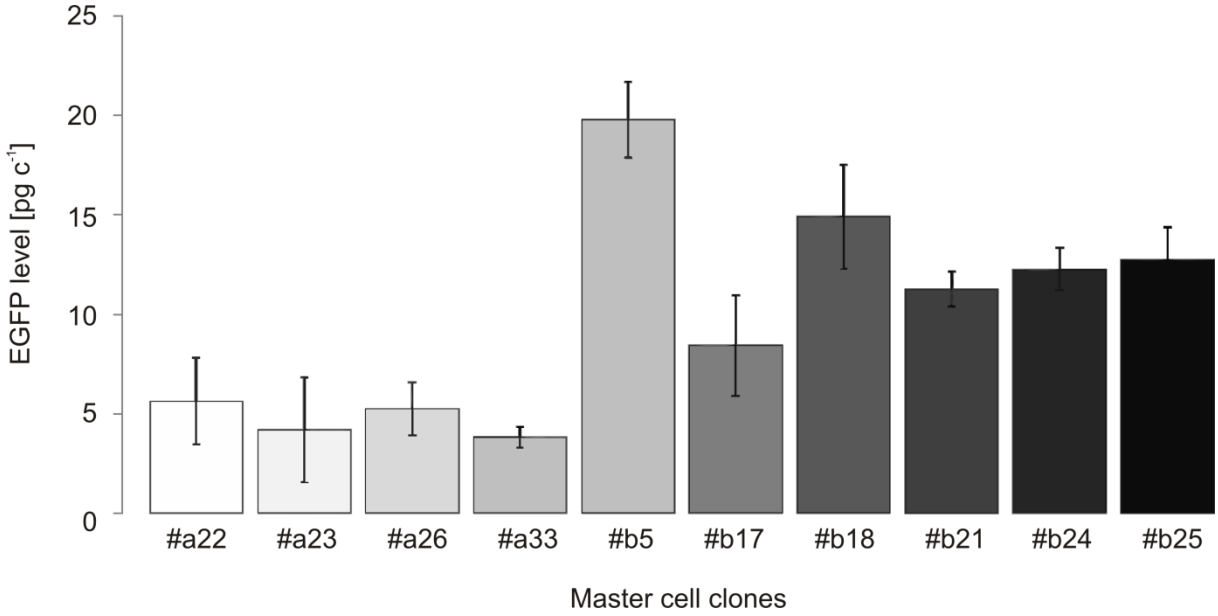


Figure 3-4: Representative GFP production level of ten master cell clones tagged by pEF-FS-EGFP-dneo (SWI3 master cell clones). The GFP concentration was determined by fluorescence spectrometry in triplicate from independent four day cultures. The concentrations were normalized to the viable cell number. A commercial GFP protein served as standard. Error bars = standard deviation.

3.2 Flp-mediated excision

The FLEx strategy facilitates cell line development without the traditional antibiotic selection. As described above, master cell clones were directly selected according to their GFP expression. In the following, the GFP gene was deleted for the expression of the genes of interest, scHGF and hLAMP3-prox. Due to the high efficiency of FLEx, no antibiotic selection was needed in this step, too.

3.2.1 scHGF cell line generated by FLEx

First, three stable GFP master cell lines (SWI4-18, SWI4-21, SWI4-25) established for scHGF with $pEF_{F3}EGFP_{F3}$scHGF were nucleofected with 5 µg pFlpO-puro encoding an optimized Flp recombinase (FlpO). FlpO is a codon improved version of Flpe, which is a variant of wild-type Flp from yeast created by molecular evolution (Buchholz *et al.*, 1998). FlpO is about five times more active than Flpe and its activity is similar to Cre in mammalian cells (Raymond & Soriano, 2007).

Upon expression in the master cells, FlpO catalysed a double-reciprocal crossover of the FRT sites resulting in reporter gene excision (Fig. 3-5 A). Thereby, the scHGF transgene was transcriptionally activated. Five days post transfection, 16%, 18% and 36% of the three transfected master clones were detected as GFP negative by flow cytometry (Fig. 3-5 B). Non-fluorescent cell clones were isolated for each of the three

master cell lines – one or two subclones per master cell line. Reporter gene excision was confirmed on the genomic level by PCR with the primers 30 and 5 (Fig. 3-5 C). Generally, two to three rounds of subcloning were necessary to obtain homogenous cell clones. Taken this into account, the cell line development from the first transfection to the pure production clone took about four months.

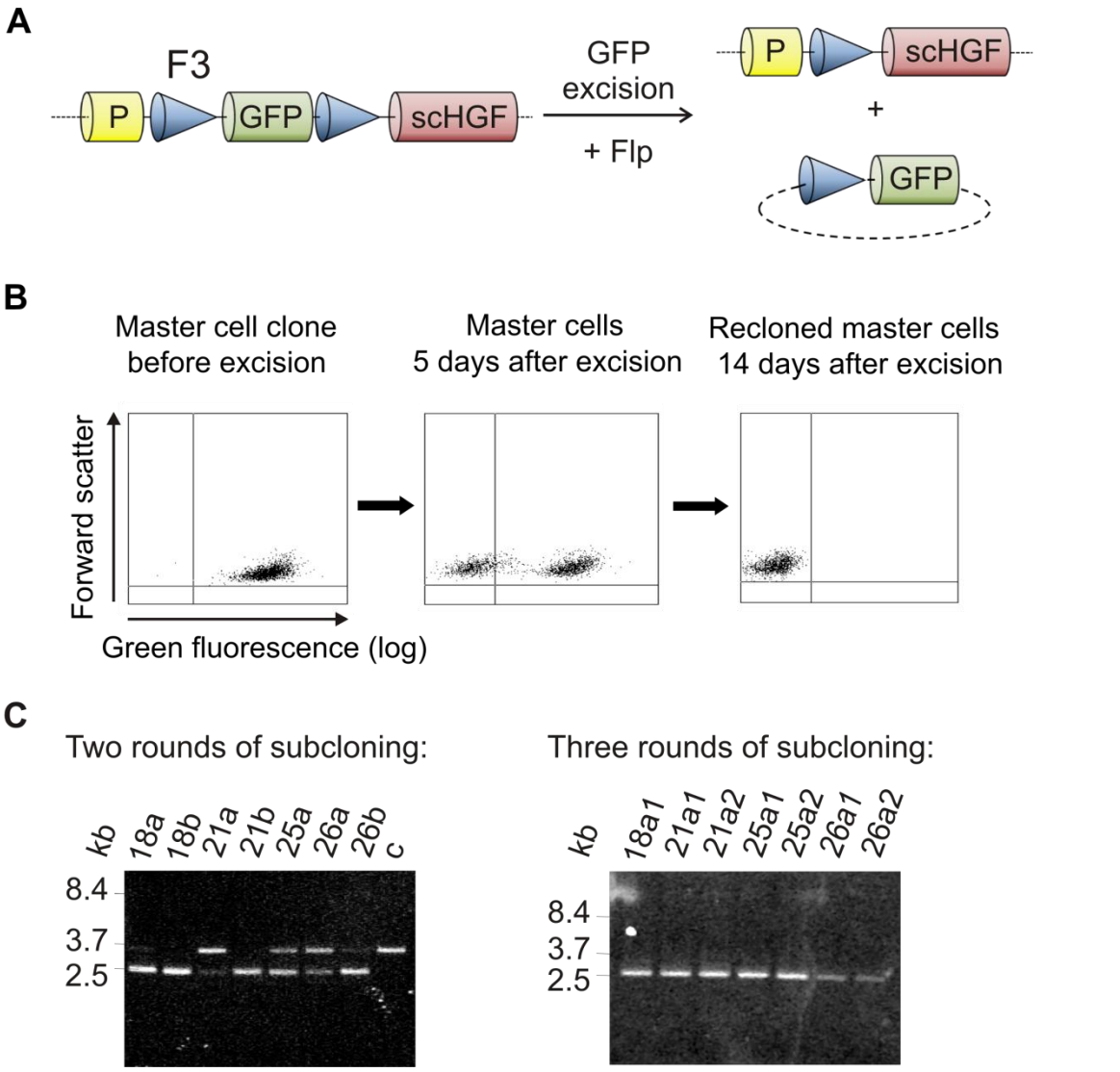


Figure 3.5: Flp-mediated reporter gene excision. (A) Schematic representation of the FLEx strategy. (B) Analytical FACS of a homogenous and stable GFP master cell clone (first plot). The cell line was transiently transfected with a Flp expression vector for GFP excision. Five days later, GFP excision led to a second, non-fluorescent cell population (second plot). Subcloning resulted in GFP-negative clonal cell lines expressing scHGF (third plot). (C) Confirmation of GFP excision by PCR. The chromosomally integrated gene cassette was amplified after two and three rounds of subcloning, respectively. The amplified gene cassette consists of 3.4 kb before and 2.4 kb after GFP excision. Cones = FRT sites (F3); P = EF-1α promoter; c = master clone number 18 before excision as negative control.

scHGF expression was detected in all tested cell lines by Western blot analysis. Subclones derived from the same master cell line expressed equal amounts of scHGF, due to their isogenicity (data not shown). Clones derived from different master clones had divergent expression levels in Western blot analysis (Fig. 3-6). The relative height of scHGF expression corresponded to the mean green fluorescence of their master cells in flow cytometry (348 a.u. for SWI4-18, 218 a.u. for SWI4-21 and 593 a.u. for SWI4-25).

The most productive scHGF cell clone SWI4-25a1 was cultivated over 11 weeks (22 passages) in a 40 mL spinner flask. Productivity and stability were analysed by ELISA after 7 passages and after 22 passages (Fig. 3-6). An average amount of 2.3 mg/L scHGF was produced in four days, and the mean specific productivity of triplicate measurements was about 1 pg per cell per day (pcd). The production was stable over this time, as product yield and specific productivity in batch culture did not change significantly (Fig. 3-6).

A production cell line for wild type human HGF/SF (Gherardi *et al.*, 2006) (provided by Ermanno Gherardi, MRC, Cambridge, UK) that had been established by conventional antibiotic selection produced significantly less under the same conditions (1.7 mg/L; 0.5 pcd; Fig. 3-6). In percent, the increase in product concentration of the FLEx clone was lower than the enhancement in specific productivity due to slower cell growth.

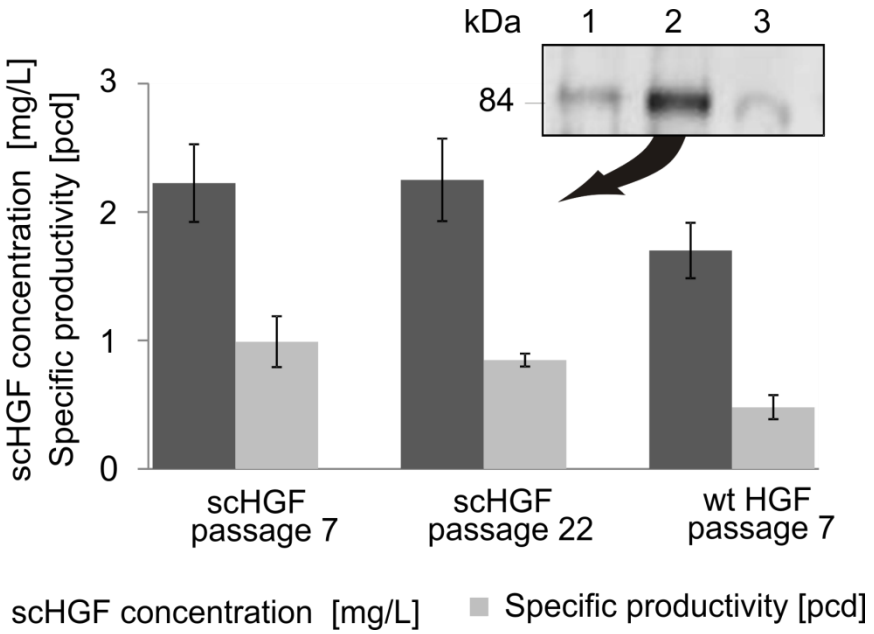


Figure 3.6: scHGF production and genetic stability. scHGF was detected by Western blot in the supernatants of stable cell lines established by cell sorting and FLEx. Lane 1: SWI4-18b, lane 2: SWI4-25a1 and lane 3: SWI4-21a. The cell clone yielding the most intense band (SWI4-25a1) was chosen for scHGF production. HGF/SF concentration in supernatants and specific productivities of stably transfected CHO Lec3.2.8.1 cell lines

were measured by ELISA. The productivity of the cell line SWI4-25a1 was compared to a conventionally established cell line for wild-type (wt) HGF/SF. scHGF production was constant over 22 passages. Dark grey = scHGF concentration, light grey = specific productivity, pcd = pg per cell per day. Error bars = standard deviation.

3.2.2 hLAMP3-prox cell line generated by FLEx

The excision of the GFP reporter gene in stable hLAMP3-prox master cell lines was performed as described for scHGF. Homogeneously non-fluorescent clonal cell lines were derived from two distinct master cell lines. Western blotting showed that three cell clones derived from one of the hLAMP3-prox master clones produced equal amounts of protein, and considerably more than a cell clone derived from the second master cell line (Fig. 3-7). The hLAMP3-prox sequence has three N-glycosylation sites. Fig. 3-7 shows that the protein is produced in different sizes with molecular mass differences corresponding to the mass of the high-mannose glycan unit, $GlcNAc_2Man_5$, of 1234 Da. This suggested that protein bearing three, two and one glycans was secreted, which was later confirmed by mass spectrometry of purified protein (see chapter 3.6.2 below).

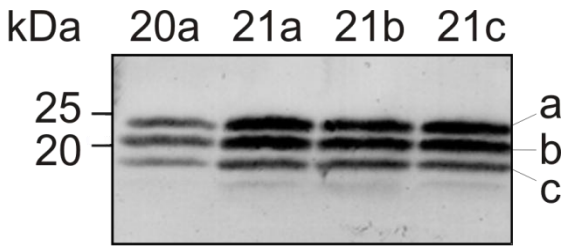


Figure 3-7: hLAMP3-prox cell lines. The hLAMP3-prox product was detected by Western Blot analysis in the supernatant of hLAMP-3-prox cell lines which were established by cell sorting and FLEx. a, b, c = hLAMP3-prox with three, two and one occupied *N*-glycosylation sites.

These results demonstrated that the excision strategy is a robust way to establish stable production cell lines. It circumvents the use of antibiotic selection which can lead to gene silencing (Liu *et al.*, 2006). Production cell lines were repeatedly established within four months. However, the master cell lines were specific for a single gene of interest. That means, novel master cell lines had to be established for each glycoprotein construct, which was regarded as inefficient. Transfected CHO Lec3.2.8.1 cells did not always survive the procedure of preparative cell sorting for unknown reasons. Therefore, the system was developed further and new master cell lines, which could be employed for arbitrary genes of interest, were established.

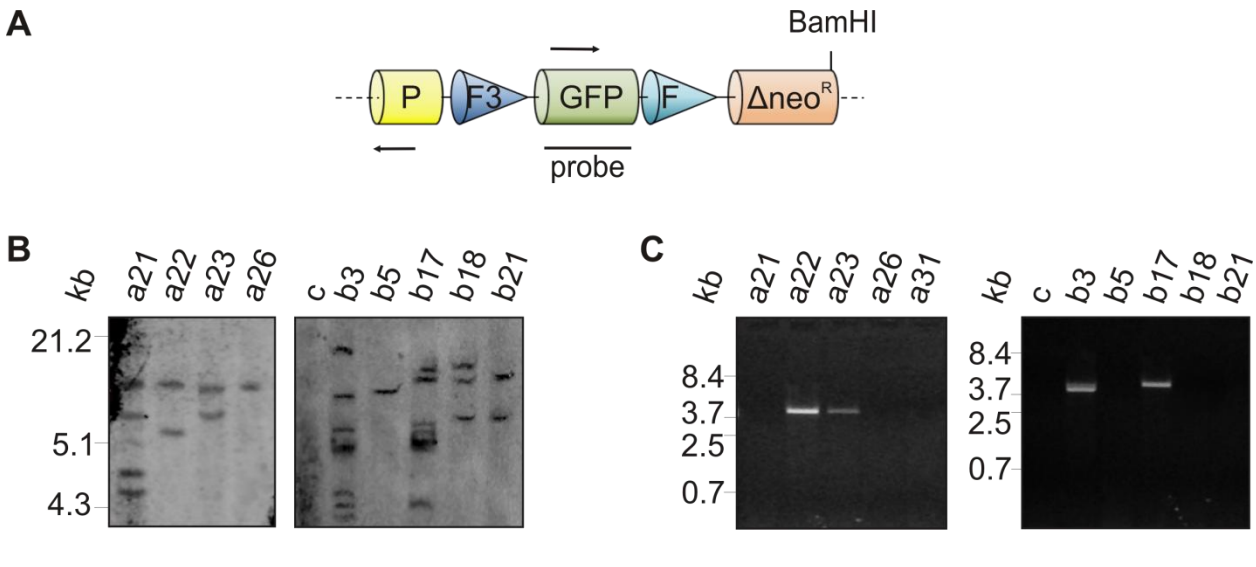
3.3 RMCE systems

Recombinase-mediated cassette exchange overcomes the limitations of the FLEx strategy described above. Once established, RMCE master cell lines can be converted to produce any protein of interest with heterologous exchange plasmids. Efficient RMCE requires master cell lines bearing only a single copy of the tagging vector, which avoids incomplete exchange and reduces the risk of illegitimate recombination leading to chromosomal rearrangements (Derouazi *et al.*, 2006) and tandem-repeat induced gene silencing (McBurney *et al.*, 2002).

3.3.1 Characterisation of the vector integration locus

Transgene copy number in the stable RMCE master cell lines described in section 3.1 was analysed by Southern blots and PCR (Fig. 3-8). Southern blots of genomic DNA digested with BamHI, which cuts downstream of the RMCE cassette, were probed for GFP. A single copy integration of the vector pEF-FS-EGFP-dneo resulted in a single band of at least 3.25 kb, depending on the integration locus. If present, further bands indicated integration of the vector at additional genomic loci (Fig. 3-8 B). Altogether, three out of 15 cell clones that were tested (SWI3a-26, SWI3b-5, SWI3b-24) were found to contain a single copy of pEF-FS-EGFP-dneo (Figure 3-8 D). For pEF-FS-EGFP, eight out of 30 contained a single copy. The presence of tandem-repeat integrations (concatemers) was analysed by PCR with the primers 22/23 (Fig. 3-8 C). Concatemer-specific PCR products were obtained from 14 multi-copy pEF-FS-EGFP cell clones (3.2 kb) and five multi-copy pEF-FS-EGFP-dneo cell clones (3.9 kb). The proportion of single-copy transgene cell line obtained by nucleofection of CHO Lec3.2.8.1 cells was quite high, about 27 % for the tagging vector pEF-FS-EGFP and 20 % for pEF-FS-EGFP-dneo, which compares well with previous studies on CHO and HEK293 cells (Nehlsen *et al.*, 2009).

To confirm completeness of the integrated transgenes, the human EF-1α promoter, plus the FRT cassette and the neomycin resistance gene were amplified by PCR with genomic DNA of the master cell clones (primer pair 4/5). In all tested 30 pEF-FS-EGFP and 15 pEF-FS-EGFP-dneo cell clones, the correct 2.5 kb and 3.3 kb fragments, respectively, were amplified (data not shown).



D

Tagging vector	Single copy cell clones	Multi copy cell clones	Concatemers in multi copy cell clones
pEF-FS-EGFP	8	22	14
pEF-FS-EGFP-dneo	3	12	5

Figure 3-8: Determination of vector copy number integrated in the CHO Lec3.2.8.1 genome. (A) Schematic representation of the tagging vector pEF-FS-EGFP-dneo. (B) Southern blot analysis of tagging vector pEF-FS-EGFP-dneo copy numbers in nine potential master cell clones. The CHO Lec3.2.8.1 host cells were included as negative control (c). Genomic DNA was digested with BamHI and a 755 bp fragment of the GFP gene was used as probe. For a single copy cell clone, one band of at least 3.2 kb was expected, while multiple bands indicated multiple copies. (C) PCR analysis of concatemers in ten potential master cell clones and the CHO Lec3.2.8.1 (c, negative control). Primers are marked by horizontal arrows in panel A. PCR products were amplified only in the presence of multiple tagging vector copies at the same integration site (D) The table summarizes the Southern blot results for cells lines tagged with pEF-FS-EGFP or pEF-FS-EGFP-dneo. Dark/light cones = F3/F FRT sites; P = promoter; Δneo^R = ATG- and promoter-deficient neomycin resistance gene.

3.3.2 Initial RMCE attempts

RMCE performance of the master cell lines tagged by pEF-FS-EGFP (SWI2 clones) was evaluated with a promoterless exchange vector encoding the red fluorescent protein (RFP). The RFP gene, lacking a promoter, is inactive until it is positioned adjacent to the EF-1α promoter present in the master cells. This promoter trap guarantees that the incoming transgene can only be expressed upon targeted integration (Qiao *et al.*, 2009; Seibler & Bode, 1997).

1.5 x 10^6 cells of four stable master clones were co-transfected in duplicate approaches with different amounts of the Flp expression vector pFlpo-puro (1, 2.5 and 4 µg) and the RFP exchange vector pFS-RFP (4, 2.5, 1 µg, respectively). According to the RMCE principle, exchange of GFP by the RFP cassette should result in the loss of GFP and

appearance of RFP fluorescence, serving as a marker to select targeted cells by FACS. The transfected cells were analysed two, three and five days after transfection by flow cytometry. However, no RFP^+, GFP^- cells were observed (data not shown), suggesting that the exchange efficiency was too low to isolate the events without antibiotic selection.

To overcome this problem, the exchange vector pFS-HygTk was constructed containing a hygromycin thymidine kinase fusion gene (HygTk) flanked by F3 and F FRT sites (Fig. 3-9 A). Thus, targeted cells gained resistance against hygromycin B. This strategy was tested with five single-copy master cell clones. 1.5×10^6 cells of each clone were co-transfected with 1 µg pFS-HygTk and 4 µg pFlpo-puro. Two to about 600 colonies appeared under 12 days of hygromycin B selection in different experiments. However, all or most of the colonies were GFP^+ in fluorescence microscopy and flow cytometry. Five to six GFP^- colonies were subcloned and expanded for each of the three master cell clones (Fig. 3-9 B). Cassette exchange was confirmed by PCR with genomic DNA (Fig. 3-9 C) and the primer pair 4/5, which amplified the FRT cassette, resulting in different sizes before and after exchange. PCR on all tested targeted genomes gave rise to 3.8 kb bands, corresponding to the exchanged cassette as expected, thus confirming successful cassette exchange.

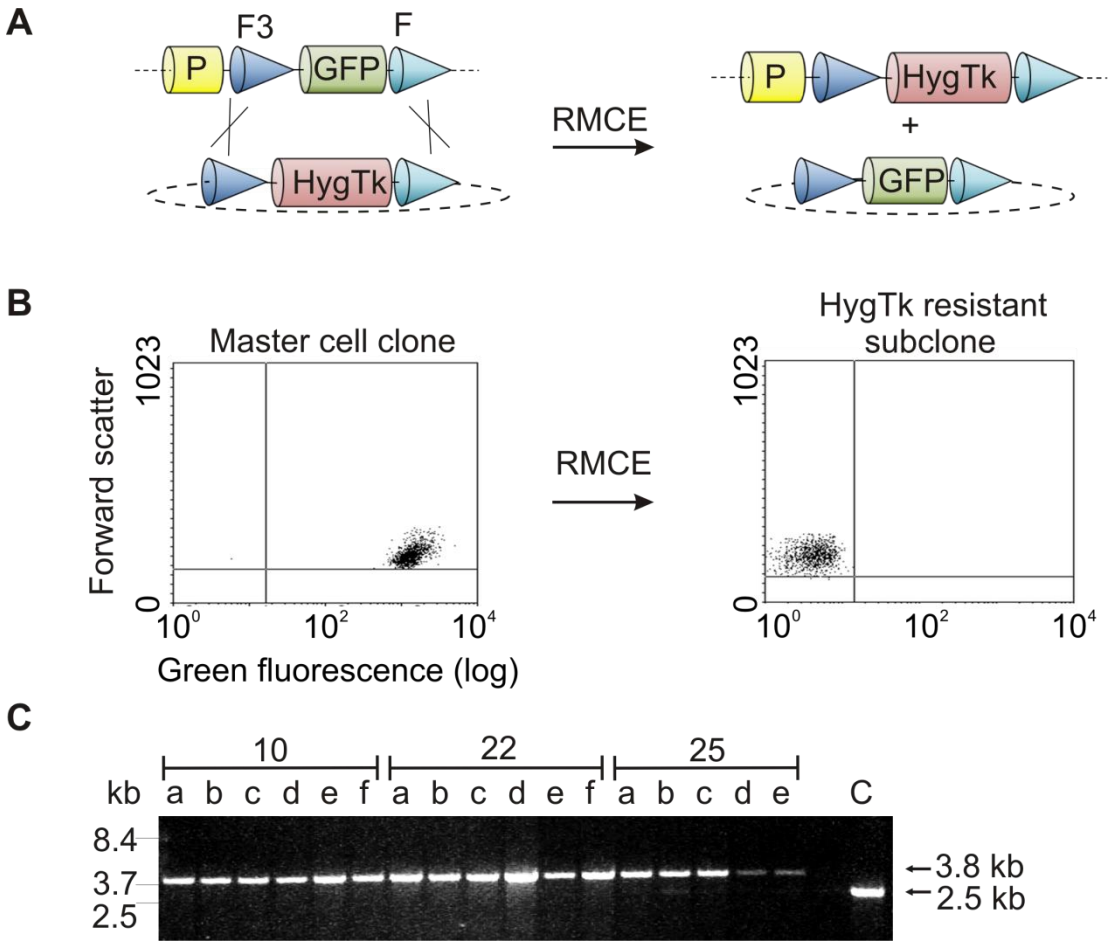


Figure 3-9: Generation of HygTk cell clones by RMCE. (A) Schematic representation of the RMCE system with promoter trap and integration of the hygromycin resistance gene fused to thymidine kinase for negative selection (HygTk). (B) Representative FACS profile of a SWI2-25 master cell clone and one subcell clone after RMCE and hygromycin B selection, demonstrating loss of green fluorescence. (C) PCR analysis of selected non-fluorescent subcell clones confirming RMCE. Six subcell clones (a-f) were analysed from master cell line SWI2-10 and -22 and five subcell clones (a-e) were analysed from master cell line SWI2-25. Dark/light cones = F3/F FRT sites; P = EF-1α promoter; c = master cells (SWI2-25) used as negative control.

Further RMCE approaches with an exchange vector bearing the scHGF gene as GOI followed by an IRES element and the HygTk gene for selection, however, failed in growing GFP^- colonies. In consequence, the HygTk strategy was not explored further. The HygTk subclones that were obtained could be used as master cells themselves. An exchange of HygTk by a GOI could be negatively selected by ganciclovir, which is a pro-drug excluding non exchanged parental cells from growing due to the viral thymidine kinase expression (Seibler *et al.*, 1998). The tymidine kinase converts ganciclovir into a phosphorylated nucleotide analog, which incorporates into the DNA of replicating eukaryotic cells and causes their death (Tomicic *et al.*, 2001).

3.3.3 RMCE with selection trap

The second RMCE strategy used a selection trap in order to enhance the stringency of colony formation. To this end, the tagging vector pEF-FS-EGFP-dneo was designed with a non-functional Δneo^R gene lacking the promoter and the start codon (Fig. 3-10 A). The Δneo^R gene is located downstream of the FRT sites and can only be activated by the incoming exchange vector upon recombination.

For exchanging GFP against RFP, nine different master cell clones were co-transfected with 1 µg pFlpo-puro and 4 µg of the RFP exchange vector pFS-RFP-PGK. Upon recombination, a PGK promoter and a start codon located on the exchange vector activated the Δneo^R gene at the tagged locus. Two weeks after transfection, G418 resistant colonies were obtained for all RMCE reactions at a frequency of about 3×10^{-5} (40-50 colonies from 1.5×10^6 transfected cells). Two to six cell clones were isolated per transfection and their fluorescence was analysed by flow cytometry (Fig. 3-10 B). Eight master clones gave rise to exclusively red fluorescent cells, while two master cell clones resulted in populations with both red and green fluorescence, indicating incomplete RMCE. The exclusively red fluorescent subclones were propagated further. RFP expression was stable over at least five weeks, as tested by flow cytometry (Fig. 3-10 C).

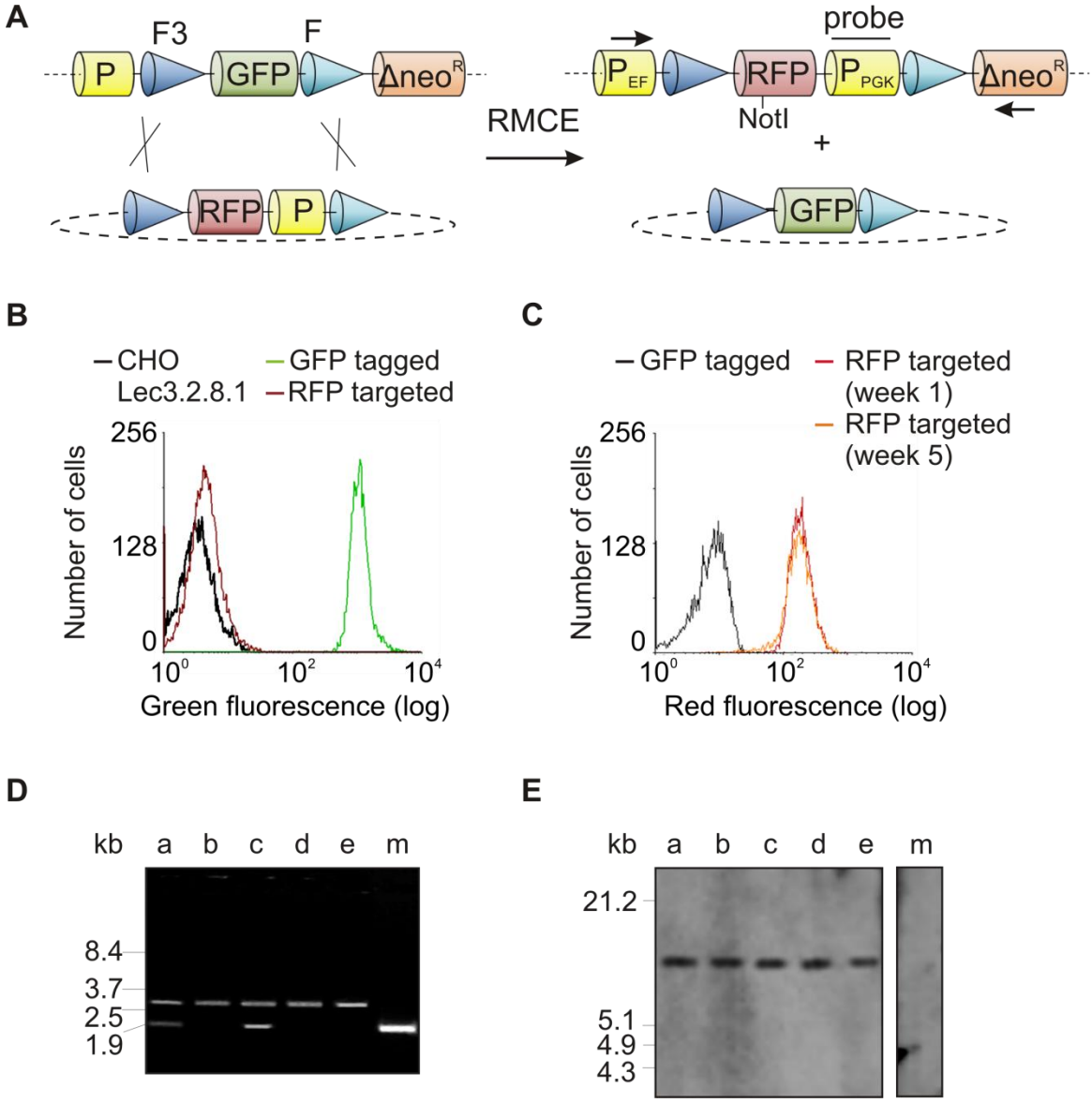


Figure 3-10: Generation of RFP production cell clones by RMCE. (A) Schematic representation of the RMCE system with promoter trap and neomycin selection trap. The reporter gene is exchanged against RFP. (B) Green fluorescence profiles comparing a representative GFP tagged master cell clone (green) with a RFP targeted cell clone (red) and CHO Lec3.2.8.1 cells (grey) as a negative control. (C) Red fluorescence profile showing the stable RFP expression of a representative targeted subcell clone after one week (orange) and 5 weeks (red) in culture. The fluorescence profiles of the RFP subcell clone were compared to that of the corresponding master cells (grey). (D) PCR verification of RMCE on genomic level. The hybridization sites of the primers are marked in panel A by horizontal arrows. They amplified the FRT site flanked gene cassette. The size difference between the amplified GFP (1.9 kb) and RFP (2.7 kb) gene cassettes enabled the identification of correctly exchanged subcell clones. Three of the five subcell clones that were tested (a-e) had a complete exchange. (E) Identification of random targeting vector integrations by Southern blot analysis. Genomic DNA was digested by NotI and a 667 bp fragment of the PGK promoter was used as probe. There were no extra bands indicating random integration of the targeting vector detected in the five RFP subcell clones (a-e). Cones = FRT sites (dark = F3, light = F), P = promoter, Δneo^R = ATG-deficient, promoterless neomycin resistance gene, m = master cell clone.

Genomic DNA was prepared from the remaining seven subclones for PCR amplification of the exchanged cassette by primers 30 and 7 (Fig. 3-10 D). The expected 2.7 kb PCR product for the exchanged gene cassette was obtained for the tested single-copy subclones SWI3a-26 and SWI3b-5, confirming complete RMCE (Table 3-1). RFP cassette exchange was also detected in the multi-copy cell clones by PCR. However, three of the six tested multi-copy master cell clones resulted in a subset of clones with incomplete exchange; both the GFP and the RFP cassettes were detected by PCR in these subsets, as representatively shown in Figure 3-10 D lane a and c. The GFP gene copies that remained in these subcells were obviously inactive as green fluorescence was not detectable. Southern blot analysis proved that the RFP exchange vector did not integrate randomly (Figure 3-10 E). Moreover, PCR with Flp-specific primers 10 and 11 did not detect random integration of the Flp expression vector for any of the subcells that were analysed (data not shown).

Table 3-1: Recombination and production properties of different master cell clones with a RFP targeting vector. Copy number of tagging vector was determined by Southern Blot analysis and concatemers were detected by PCR. GFP concentration was measured from cell extracts of four day cultures by fluorescence spectrometry. The results were correlated to a standard curve of commercial GFP fluorescence intensities of known concentration. conc. = concentration.

Master cell clones	a-22	a-23	a-26	a-33	b-5	b-18	b-25
Transgene copy number	2	2	1	3	1	3	2
Concatemers	yes	yes	no	no	no	no	no
GFP conc. [mg/ L]	19.2	13.7	12.2	11.2	19.9	29.8	20.5
Proportion of clones retaining GFP copies	6/6	3/5	3/3	0/6	2/2	2/2	3/3

Establishing and analysing the RFP subclones from transfection to the first cryo-conservation took seven weeks (Fig. 3-11). Upon transfection, the cells grew 3 weeks adherently in the presence of G418 until cell colonies were isolated. To ensure that the isolated colonies originated from a single cell, subcloning by serial dilution was performed. The single cell clones were then expanded by cultivation in suspension in 6 well plates. These cells were the basis for flow cytometry, PCR, Western blot analyses and cryo-preservation.

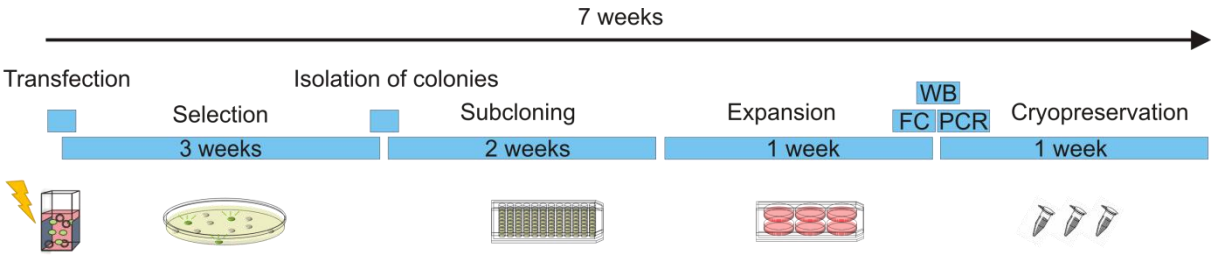


Figure 3-11: Timeline for RFP cell line development by RMCE. Master cells were co-transfected with Flp-expression and targeting vectors. Positively exchanged cell clones were selected by G418. Two weeks after transfection, RFP positive colonies were isolated and subcloned by serial dilution. Single cell clones were expanded to grow in 6 well plates. To confirm the cassette exchange, the cells were analysed by flow cytometry and PCR analysis. The recombinant protein production was controlled by Western Blot analysis before cryopreservation of the cell clones. It took 7 weeks from transfection of the master cells to cryopreservation of the first aliquot of RFP cell clones. More time may be required, e.g., if an additional round of subcloning is required to obtain homogenous cell clones. WB = Western Blot, FC = flow cytometry.

3.3.4 Evaluation of RFP production

RMCE performance, expression strength and genetic homogeneity of subclones are important features of a good master cell line. Recombinant protein production of seven GFP expressing master cell lines and two to four corresponding RFP subcell clones was quantified (Fig. 3-12). The GFP productivity of master cells from four day cultures varied between 4 pg/ cell and 20 pg/ cell according to fluorescence spectroscopy (Fig. 3-4). The expression level distribution corresponded to flow cytometry intensities (Fig. 3-12).

As expected from isogenic clones, subclones obtained by targeting the same master cell line produced RFP homogenously (Fig. 3-12). The two master cell lines containing transgene concatemers, SWI3a-22 and SWI3a-23, gave rise to cell lines with notably low RFP expression, which can be explained by the reduction of the concatemers to single copies during the recombination reaction. Absence of concatemers upon RMCE was confirmed by PCR with primers 9 and 12 (data not shown). RMCE with the remaining five master cell lines, although producing different amounts of GFP, resulted in subcell clones expressing a similar, saturating level of RFP. Thus, factors other than transcriptional activity might limit the maximal RFP concentration in this cell type.

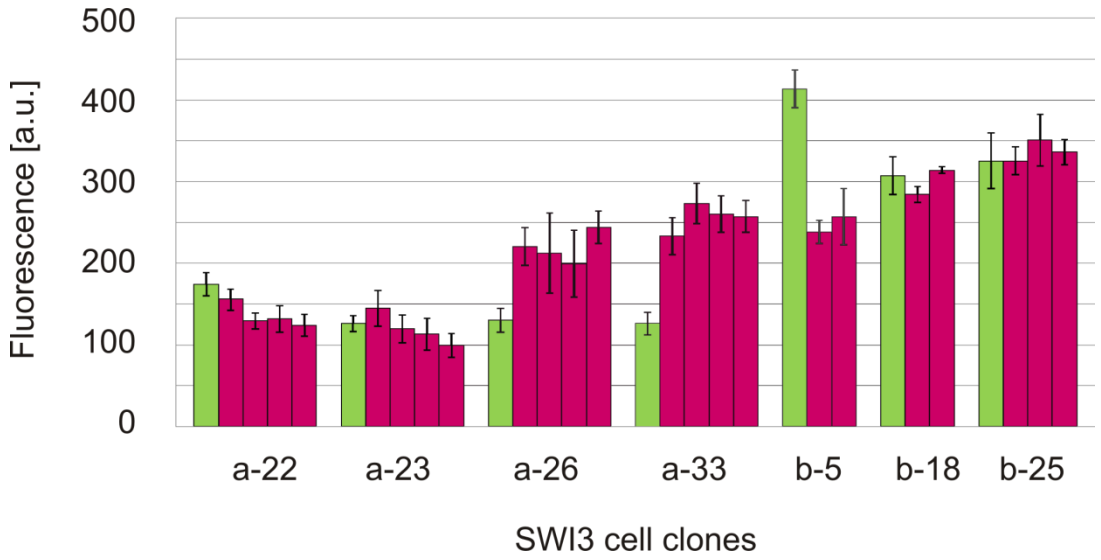


Figure 3-12: GFP and RFP production levels. The fluorescence strength of seven GFP master cell clones (green) and the corresponding RFP targeted subcell clones (red) were measured by flow cytometry. Each cell clone was analysed in triplicate. Error bars = standard deviation. GFP fluorescence was scaled down by a factor of 10.

Further experiments were mainly performed with the master cell clone SWI3a-26, which was genetically stable, carried a single transgene copy of high transcriptional activity and performed well in RMCE with 100% recombinant and highly fluorescent subclones.

3.4 RMCE production cell lines

To demonstrate the feasibility of the RMCE approach for producing protein for crystallography, production cell lines for nine different target proteins were established. Experimental work of this part was supported by Sarah Torkaski, Agathe Duda, Jörn Josewski and Lilia Polle. The proteins which are listed in Table 3-1 include glycoproteins, a GPI-anchored protein and an intracellular protein. The different LAMP constructs were chosen upon transient expression tests in HEK293 and CHO Lec3.2.8.1 cells (performed by Dr. Konrad Büssow).

Table 3-2: Recombinant proteins produced by RMCE derived from CHO Lec3.2.8.1 cells. The molecular weight was calculated from the amino acid sequence without signal peptides, glycosylation and GPI-anchors. MW = molecular weight, Glyc. sites = N-glycosylation sites as indicated by the sequon Asn-X-Ser/Thr in the amino acid sequence.

Symbol	Construct	Protein name	Species	Localisation	MW (kDa)	Glyc. sites
FcεRIα	ectodomain	High Affinity IgE receptor α	Homo sapiens	secreted	21	7
GILT	Full length	Gamma-interferon inducible lysosomal thiol reductase	Homo sapiens	secreted	22	3
mLAMP2-dist	Membrane-distal luminal d.	Lysosome-associated membrane protein 2	Mus musculus	secreted	19	7
mLAMP2-lum	Luminal domain	Lysosomal-associated membrane protein 2	Mus musculus	secreted	40	16
hLAMP2-lum	Luminal domain	Lysosomal-associated membrane protein 2	Homo sapiens	secreted	39	7
rLAMP2-dist	Membrane-distal luminal d.	Lysosomal-associated membrane protein 2	Rattus norwegicus	secreted	19	8
hLAMP3-prox	Membrane-proximal luminal d.	Lysosomal-associated membrane protein 3	Homo sapiens	secreted	19	3
NALP3	Full length	NACHT, LRR and PYD domains containing protein 3	Homo sapiens	intracellular	118	0
scHGF	Single-chain variant	Hepatocyte growth factor/ Scatter factor	Homo sapiens	secreted	84	4
shPrP	Full length	Prion protein	Meso-cricetus auratus	Secreted, GPI-anchored	24	2

3.4.1 Pilot expression test for LAMP constructs

Proteins of the LAMP family are heavily glycosylated type I membrane proteins with a short cytoplasmic C-terminal domain. Their N-terminal region is localised in the lysosomal lumen. It is composed of two similar domains, which are connected by a proline rich, flexible hinge sequence. The aim of the pilot expression test was the identification of protein constructs with minimal flexibility for optimal crystal formation on the one hand and good protein yield on the other hand.

A set of 33 LAMP domains of different species was cloned into the mammalian expression vector pEFFS-sigHA with a C-terminal HA tag. HEK293 and CHO

Lec3.2.8.1 were transiently transfected and secreted proteins were detected by Western blot (data not shown). Five proteins were secreted with relatively homogenous size and high yield and were selected for construction of stable cell lines: the complete luminal region of human and mouse LAMP-2 (hLAMP2-lum, mLAMP2-lum), the membrane-distal domain of mouse and rat LAMP-2 (mLAMP2-dist, rLAMP2-dist) and the membrane-proximal domain of human LAMP-3/DC-LAMP (hLAMP3-prox).

3.4.2 Generation of production cell lines

RMCE for the production cell line development was generally performed as described for RFP. Adding selection pressure five rather than two days post transfection improved recombination frequency for the new constructs. Depending on the construct, the optimal mass ratio of co-transfected targeting vector and Flp-expression vector was 1:4, 2:3 and 1:1, corresponding to a molar ratio of approximately 1:2.

Stable production cell lines were derived from SWI3a-26 for scHGF, shPrP^{c}, hLAMP2-lum, mLAMP2-lum, FcεRIα, GILT and NALP3 and RMCE was confirmed by PCR in 38 of 40 analysed subcell clones (data not shown). Random integration of the targeting vector could not be visualized in 10 subclones expressing scHGF by Southern Blot analysis (data not shown). This result was regarded as representative. Thus, this analysis was not performed for the other subclones. In addition, cell lines for hLAMP3-prox, mLAMP2-dist and rLAMP2-dist were established with SWI3a-33. This master cell line bore three GFP gene cassettes and the cassette exchange was proved to be incomplete in all 12 analysed cell clones, though no green fluorescence was detected and the target proteins were found in the culture supernatant by Western blot analysis (Fig. 3-13).

At least four subclones of average cell size were expanded for each construct with the exception of mLAMP2-dist where only one subclone was available. Recombinant protein production by targeted cell clones was analysed by Western blot (Fig. 3-13). Subclones derived from the same RMCE reaction expressed their transgene at a similar level, as expected for isogenic cells. With the exception of NALP3 cell clones, recombinant protein yield appeared adequate for producing sufficient amounts for crystallization. NALP3 is an intracellular protein that was solubly produced, but barely detectable by an anti-His antibody (Fig. 3-13) and by an anti-NALP3 antibody (data not shown) as well.

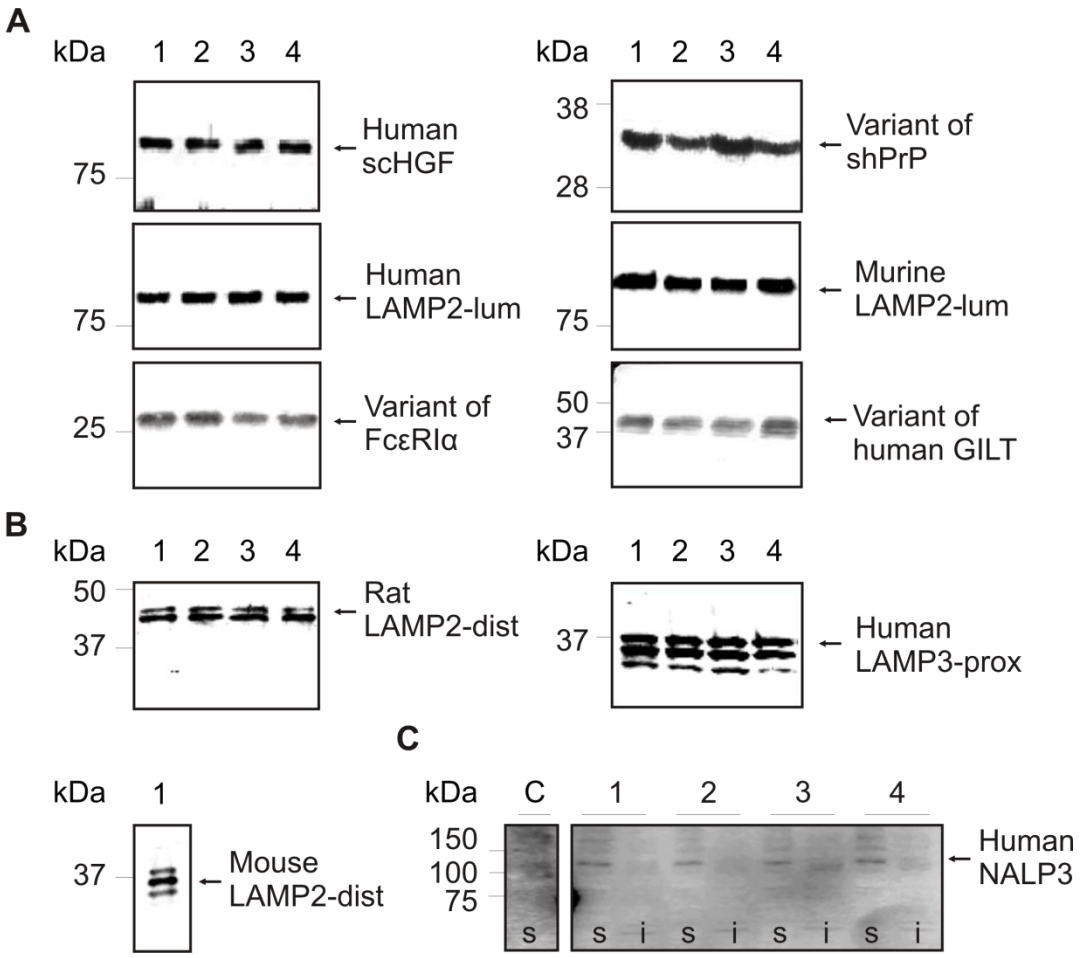


Figure 3-13: Expression test of a variety of production cell lines established by cassette exchange. For all recombinant products, four representative, isogenic subcell clones were analyzed by Western blot. (A) Analysed supernatants or cell lysates containing human scHGF, a variant of shPrP, human LAMP3-lum, murine LAMP2-lum, FcεRIα ectodomain and a variant of GILT are illustrated. All subcell clones were derived from master cell clone SWI3a-26. (B) Supernatants of cell clones producing rat LAMP2-dist, human LAMP3-prox and mouse LAMP2-dist are illustrated. These cell lines were derived from the SWI3a-33 master cell line. (C) Cell lysates of four NALP3 targeted subcell clones and the soluble lysate fraction of SWI3a-26 cells are illustrated. The subcells derived from master cell line SWI3a-26. Small amounts of recombinant human NALP3 were detected in all four soluble lysate fractions of the targeted cells and not in the negative control. C = negative control (SWI3a-26 master cells), i = insoluble, s = soluble.

3.4.3 Characterisation of scHGF subcell clones

During the cultivation of five scHGF subcell clones, strikingly different growth curves and maximal cell densities were observed, while cell viability was generally high and glucose and lactate concentrations were not limiting (Fig. 3-14 A). Productivity of scHGF per cell correlated inversely with the maximal cell density - cell lines with weak growth produced more protein (Fig. 3-14 B). The scHGF product titres among the subclones varied between 1.4 mg/L and 4.6 mg/L. Cell size measurements revealed differences

between the cell lines that correlated with the divergent cell densities and productivities (Fig. 3-14 C). Some of the subclones had increased their size during RMCE and produced more protein than normal sized cells. This increase in size correlated with a higher content of chromosomal DNA (DNA index, Fig. 3-14 B), indicating chromosomal abnormalities like aneuploidy caused by the instability of the CHO genome (Barnes *et al.*, 2003; Huang *et al.*, 2007). The DNA index was determined by flow cytometry cell cycle analysis. The described heterogeneity for scHGF subclones was not observed in any of the other targets' cell lines.

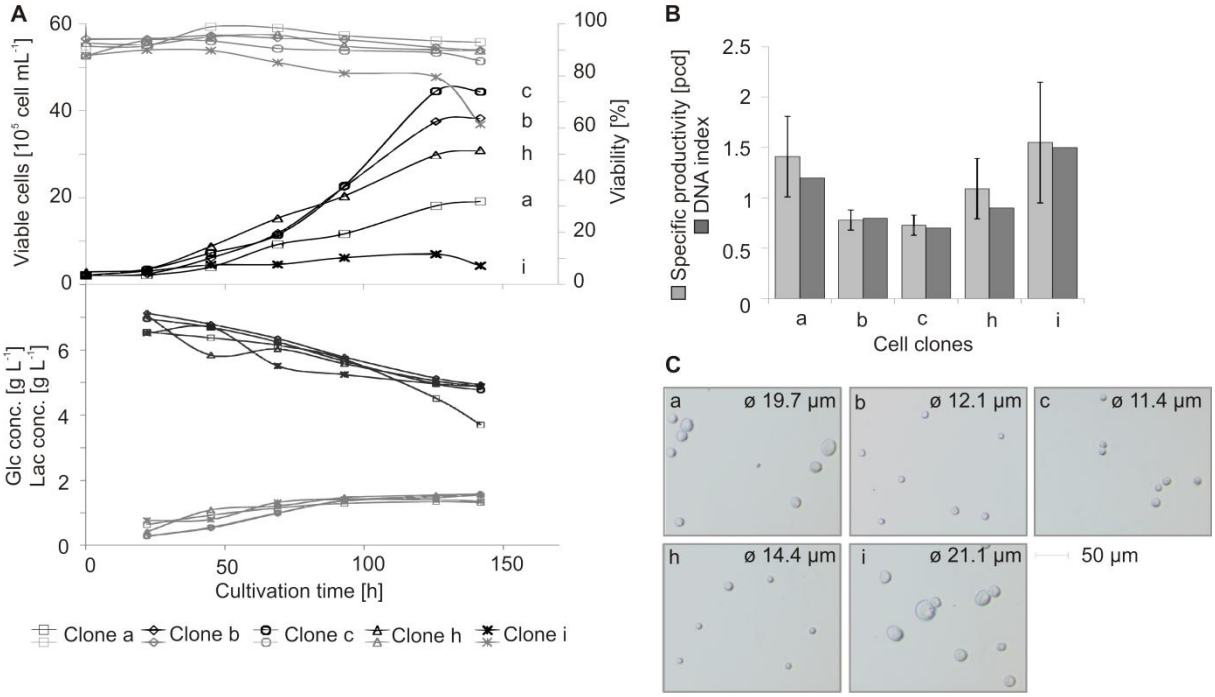


Figure 3-14: Evaluation of growth and production properties of targeted scHGF subcell clones from one master cell line. Five targeted subclones of master cell line SWI3a-26 were compared in parallel batch cultivations. (A) Course of viable cell concentrations (black), cell viabilities (grey), glucose concentration (black) and lactate concentrations (grey) in the medium of five parallel cultured scHGF subclones. (B) Productivity and DNA index of the targeted subclones (a, b, c, h, i). Secreted scHGF in the medium was measured in mg/L and normalized to the viable cell number. This resulted in the specific productivity of the cells presented in pg per cell per day (pcd, light grey). The DNA index was determined by cell cycle measurements. To this end, the cells were fixed and their DNA was stained by propidium iodide which was followed by flow cytometry analysis. (C) Cell size and morphology of the five subclones determined by phase contrast images (Axiovert100 fluorescent microscope, Carl Zeiss, Göttingen, Germany). Average cell sizes are indicated.

A clone with high productivity and acceptable growth was chosen for further analysis (clone a). Its productivity in 40 ml spinner flasks (1.4 ± 0.4 pcd; 4.3 mg/L), compared favourably to a scHGF cell clone derived from the FLEx system (SWI4-25a;

1.0 ± 0.2 pcd; 2.3 mg/L) and a conventionally established cell clone for wild-type HGF (EGT92/A20; 0.5 ± 0.1 pcd; 1.7 mg/L) which were described in chapter 3.2.1.

3.4.4 Comparison of different scHGF cell lines in perfusion processes

For structural studies, especially X-ray crystallography, milligram amounts of protein are required. Hence, the cultivation of targeted subcell clones was scaled up to 2.5 L perfusion bioreactors. Conditioned medium was continuously conducted from the bioreactor under cell retention, while fresh medium was added, until 20 L medium were perfused.

To compare the cultivation and productivity of scHGF cell lines established by RMCE and FLEx (clones SWI3a-26a and SWI4-25a1, respectively), two 2.5 L perfusion processes were performed in parallel. Bioreactors were inoculated with 0.2 x 10^6 cells/mL and cultivated for 5 days in batch modus to viable cell densities of 2.6 x 10^6 mL^{-1} (SWI3a-26a) and 4.4 x 10^6 mL^{-1} (SWI4-25a1) (Fig. 3-15). The corresponding product concentrations were 3.2 mg/L and 1.0 mg/L for SWI3a-26a and SWI4-25a1, respectively. Though having a lower cell density, the RMCE cell clone produced more scHGF than the reporter gene excision clone. As described in chapter 3.4.3, this effect again correlates with the different cell sizes of the clones. The RMCE cells had an average size of 19.4 µm in diameter, while the FLEx cells were 13.5 µm in diameter.

The cultivation temperature was shifted from 37°C to 32°C during the process (Fig. 3-15). It had been reported that the decrease in temperature prolongs culture viability and thus lengthens production phase resulting in higher product yield (Kaufmann *et al.*, 1999; Moore *et al.*, 1997). Moreover, the temperature shift may also increase the specific productivity, but this effect is variable among different recombinant CHO cell lines (Bollati-Fogolin *et al.*, 2005; Kaufmann *et al.*, 1999).

After 10 days, 22.5 L conditioned medium were produced for both processes. The peak cell densities were 8.3 x 10^6 mL^{-1} for SWI3a-26a and 12 x 10^6 mL^{-1} for SWI4-25a1 (Fig. 3-15). The final titre in the perfused culture supernatant was 1.44 mg/L and 0.9 mg/L for SWI3a-26a and SWI4-25a1, respectively.

In conclusion, the final titre of the scHGF RMCE cell line SWI3a-26a outperformed the reporter gene excision cell clone SWI4-25a1, although the latter reached a higher cell density. Since the clone SWI4-25a1 was established earlier, it was used for producing scHGF for the studies described in the following.

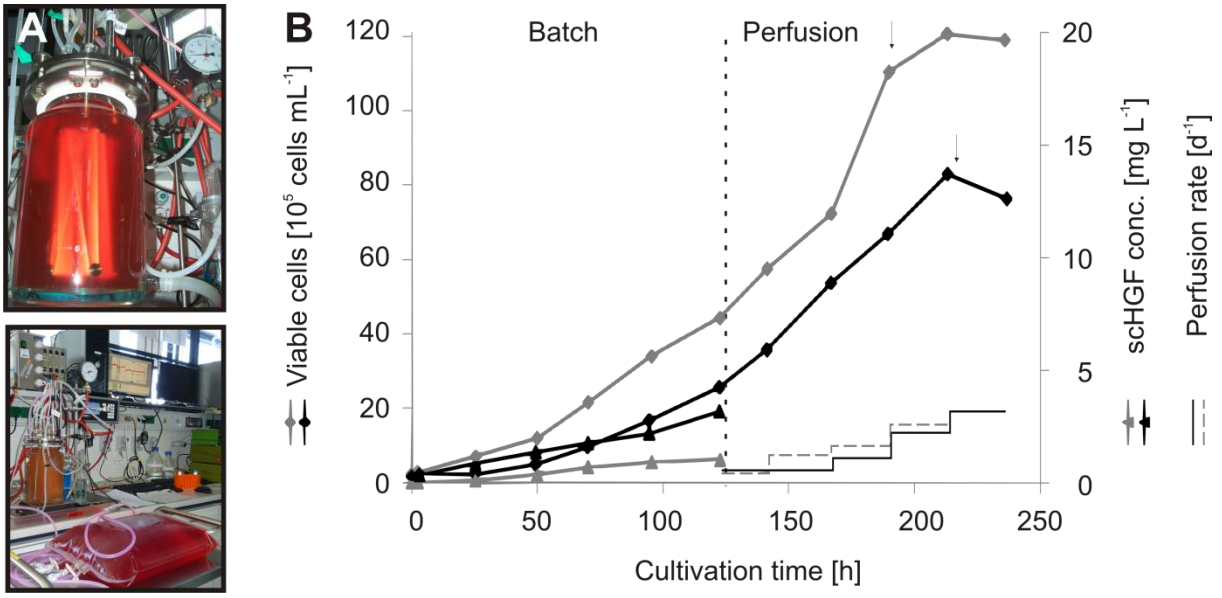


Figure 3-15: Comparison of two scHGF cell lines in 22.5 L perfusion bioprocesses. (A) Photo of a 2.5 L bioreactor cultivating the cell clone SWI3a-26a. The upper panel shows the bioreactor at process start with a cell density of 5 x 10^5 cells mL^{-1} (upper panel) and the lower panel shows the bioreactor at the end of the process with a cell density of 83 x 10^5 cells mL^{-1}. 20 L perfused medium are located in the front. (B) Course of viable cell density (rhombus), scHGF concentration (triangle) and perfusion rate (line) are shown for the parallel cultivation of cell clones SWI3a-26a (RMCE, black, line) and SWI4-25a1 (FLEx, grey, dotted line). The vertical dotted line marks the start of perfusion and the arrows mark a temperature shift from 37°C to 32°C.

3.5 Single-chain variant of HGF/SF

Natural HGF/SF is secreted as a biologically inactive single chain (sc) precursor with four N-glycosylation sites and one O-glycosylation site. It is proteolytically cleaved into the mature, biologically active two chain form (Birchmeier *et al.*, 2003). Stable CHO Lec production cell lines for wild-type HGF/SF and the ligand binding domain of its receptor Met have already been established (Gherardi *et al.*, 2006). This work comprises the production and glycosylation analysis of a non-cleavable HGF/SF mutant (scHGF) bearing two mutations (K491D and R494E) in its natural cleavage site and a C-terminally fused glycine followed by a hexa-histidine affinity tag. Table 3-3 summarizes important physico-chemical parameters of the scHGF construct.

Table 3-3: Physico-chemical parameters of human scHGF with C-terminal His_6-tag as calculated with VectorNTI (Invitrogen).

Parameter	scHGF
Length (aa)	704
Molecular weight (kDa) [a]	80.5
Molar extinction coefficient	148,300
Isoelectric point (pI)	7.9

[a] Molecular weight was calculated without signal peptide and glycosylation.

3.5.1 Production and purification of scHGF

scHGF was produced in 2.5 L batch cultivations of the CHO Lec3.2.8.1 clone SWI4-25a1. The protein was purified from the cultivation supernatant by heparin and ion exchange chromatography according to the purification protocols for natural HGF/SF by Dr. Jörn Kraußе (Fig. 3-16). Typical yields amounted to 1 to 3 mg pure scHGF per litre of culture. scHGF was mainly obtained as a single protein chain due to the mutations at the protease cleavage site. However, small amounts of the cleavage products α- and β-chain were detected by SDS-PAGE (Fig. 3-16) despite these mutations. This was confirmed by mass spectrometry and N-terminal sequencing which identified the natural N-terminus of the β-chain and thus the same cleavage position as in wt HGF between Arg494 and Val495 (Naldini *et al.*, 1992). Adding 0.1 mM pefabloc SC plus 8 (Roche) to cultivation seems to reduce the cleavage products in Western blot analysis of cultivation supernatants, but not after purification (data not shown).

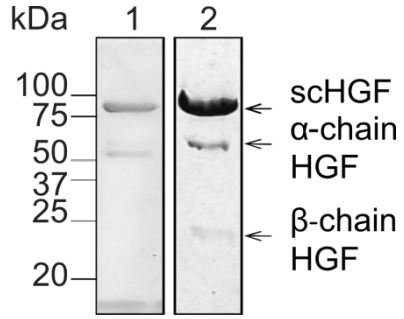


Fig. 3-16: scHGF purification. scHGF was produced by a 2.5-L batch cultivation and purified with a heparin column (lane 1), followed by ion exchange chromatography (lane 2). Proteins were analyzed by 12% SDS-PAGE and Coomassie staining.

3.5.2 Glycosylation analysis of scHGF

The heterogeneous and flexible nature of glycan modifications generally influences the crystallisation of glycoproteins (Butters *et al.*, 1999). To increase the chances of crystallisation, purified scHGF was deglycosylated with endoglycosidase H (endo H). scHGF was treated with different amounts of this enzyme over night (Fig. 3-17). Since no significant mass shift was detectable by SDS-PAGE, masses of the protein were measured before and after deglycosylation with 30 U/µg endo H at pH 5.2 by MALDI-

TOF mass spectrometry (Fig. 3-17 C), which confirmed the calculated mass shift of threefold glycosylated scHGF upon deglycosylation (84,129 to 81,090). Additional ESI-TOF mass spectrometry of tryptic peptides indicated that position N402 was unglycosylated, while the others (N294, N566, and N653) bore $GlcNAc_2Man_5$ chains.

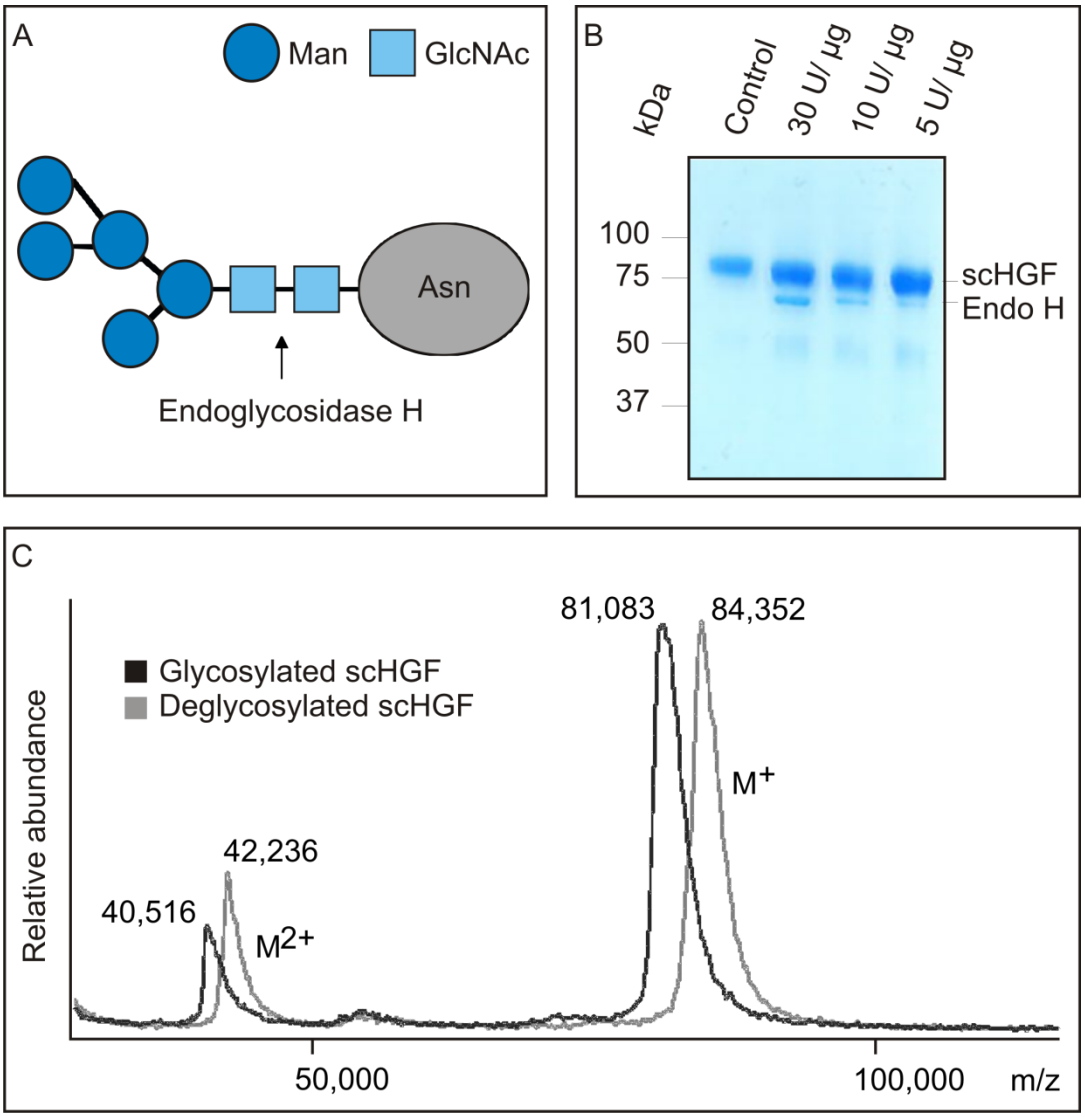


Figure 3-17: Deglycosylation of scHGF. (A) Schematic representation of a high-mannose type N-glycan as it is synthesized by the glycosylation mutant CHO Lec3.2.8.1 cells. The cleavage site of endoglycosidase H (endo H) is indicated by an arrow. (B) Endo H digestion of purified scHGF produced in CHO Lec3.2.8.1. Different endo H concentrations (5 U/µg, 10 U/µg, 30 U/µg) were tested in over night incubations and analysed by a 12% SDS gel. No significant size differences were observed between the glycosylated and deglycosylated scHGF. (C) Confirmation of enzymatic deglycosylation of scHGF by MALDI mass spectrometry. Molecular weights of 84,352 and 81,083 Da, respectively, were measured for glycosylated and deglycosylated scHGF. m/z = ion mass/ ion charge; $M^{+/2+}$ = molecular ions with one/two positive charges; Man = mannose; GlcNAc = N-acetylglucosamine; Asn = asparagine; Control = scHGF without endo H.

3.6 hLAMP3-prox

In this study, the structure of the luminal membrane-proximal domain of LAMP-3 (hLAMP3-prox) consisting of the amino acids 222 to 381 was investigated. To this end, the domain was fused with a mouse immunoglobulin signal peptide (Swissprot P01750) and a C-terminally hexa-histidine affinity tag. It was recombinantly produced with CHO Lec3.2.8.1 cell lines (Fig. 3-18). hLAMP3-prox bears three N-glycosylation sequons and two disulphide bonds predicted by homology. After establishing a chromatographic purification protocol, as detailed below, the protein was deglycosylated and crystallised, resulting in first insights into the previously unknown structure of LAMP-3/DC-LAMP and the whole LAMP family. Table 3-4 summarizes physico-chemical parameters of the hLAMP3-prox construct.

Table 3-4: Physico-chemical parameters of human hLAMP3-prox as calculated with VectorNTI (Invitrogen).

Parameter	hLAMP3-prox
Length (aa)	166
Molecular weight (kDa)[a]	19
Molar extinction coefficient	9440
Isoelectric point (pI)	5.7

[a] Molecular weight is calculated without signal peptide and glycosylation.

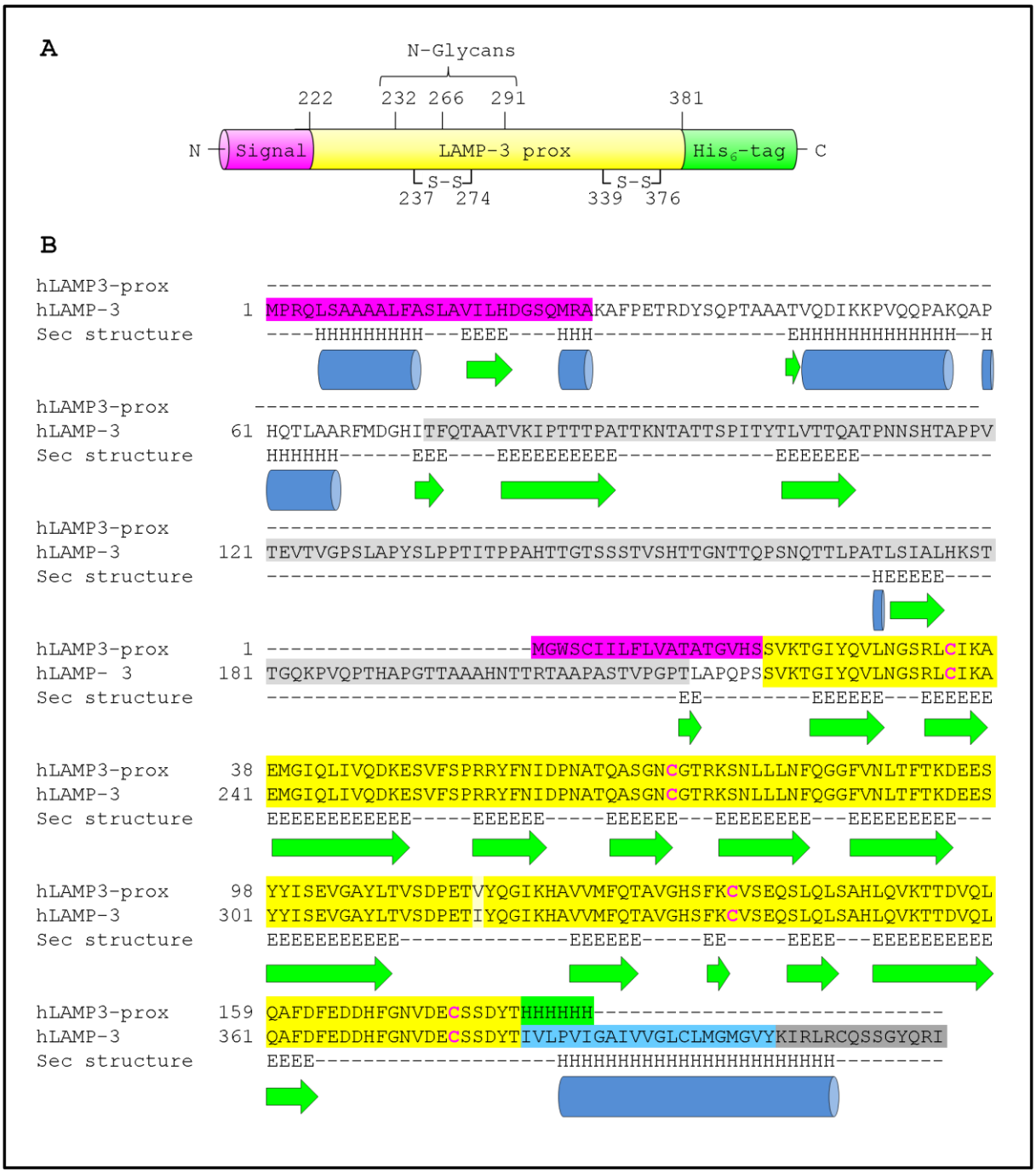


Figure 3-18: Human hLAMP3-prox construct. (A) Schematic representation of human hLAMP3-prox (amino acids 222–381 of LAMP-3, yellow) which is expressed with a mouse immunoglobulin signal peptide (Swissprot P01750, pink) and a His_6-tag (green). Positions of potential N-glycans and cysteines involved in disulphide bridges are indicated. (B) ClustalW alignment of hLAMP3-prox and the full-length sequence (hLAMP-3). The secondary structure prediction was performed by JPred (Cole *et al.*, 2008). The sequence annotation is according to Swiss-Prot entry Q9UQV4. E = strand (arrow); H = helix (cylinder). Colour code: signal sequence = pink; threonine rich region = light grey; hLAMP3-prox domain = yellow; His_6-tag = green; transmembrane region = blue; cytoplasmic domain = dark grey. Cysteines in the hLAMP3-prox domain are marked pink and a natural variation between hLAMP3-prox and hLAMP-3 is highlighted at position 318. Numbers represent amino acid positions.

3.6.1 Production of hLAMP3-prox

To evaluate the growth and production of two hLAMP3-prox-expressing cell clones established by FLEx and RMCE, parallel cultivations were performed in 2.5 L perfusion bioreactors similar to the scHGF processes described in 3.4.4.

The bioreactors were inoculated with 0.4 x 10^6 cells/ mL and cultivated for 120 h in batch modus (Fig. 3-19). Within this time, both cell clones reached a viable cell density of 3.3 x 10^6 mL^{-1} and similar maximal growth rates of about 0.4 d^{-1}. The perfusion rate was stepwise increased to one reactor volume per day (d^{-1}) after 120 h of cultivation. The cultivation temperature was decreased from 37°C to 32°C after 260 h, followed by decreasing the perfusion rate to 0.5 d^{-1}. Maximal cell densities were 8.9 x 10^6 mL^{-1} for the FLEx cell clone and 7.8 x 10^6 mL^{-1} for the RMCE cell clone, measured when the cultivations were stopped after 401 h and 429 h, respectively. Thus, the growth behaviour of the cell lines was similar in perfusion processes. However, the viability of the reporter gene excision clone dropped markedly in perfusion mode. For determining the hLAMP3-prox yield, the protein was purified from the culture supernatant.

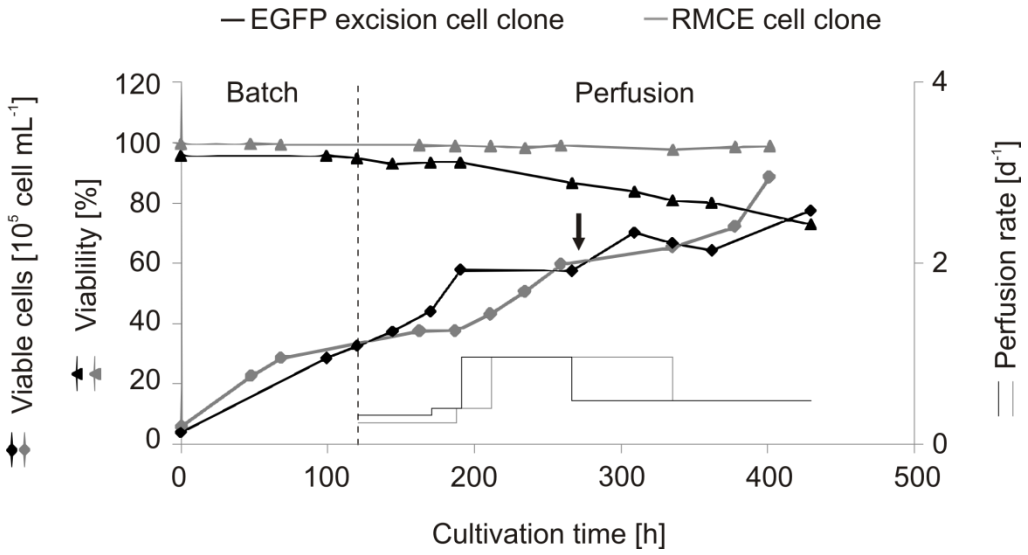


Figure 3-19: Comparison of two hLAMP3-prox cell lines in 2.5 L bioreactors. Course of viable cell density (rhombus), viability (triangle) and perfusion rate (line) are shown for the parallel cultivation of LAMP-3 cell clones derived from RMCE (grey) and FLEx (black). The vertical dotted line marks the start of perfusion and the arrow marks a temperature shift from 37°C to 32°C.

3.6.2 Purification of hLAMP3-prox

hLAMP3-prox was purified from 22.5 L conditioned medium comprising 20 L perfused medium and 2.5 L culture supernatant from SWI3a-26a and SWI4-25a cells, respectively. The supernatant was separated from the cells by centrifugation and the

total medium was concentrated to 1 L by diafiltration with PBS, followed by purification of hLAMP3-prox by immobilized metal ion affinity chromatography (IMAC). A typical chromatogram is shown in Figure 3-20 A. Captured protein eluted in two separate peaks. The first peak predominantly contained unspecifically bound protein. The second peak contained three proteins with sizes corresponding to heterogeneously glycosylated hLAMP3-prox (20 – 25 kDa). Additional proteins were still present. Since all unspecific bound proteins were larger than hLAMP3-prox, a second purification step was performed by preparative gel permeation chromatography (GPC, Fig. 3-20 B). To this end, the fractions of the second IMAC peak were concentrated to 10 mL and subjected to GPC in 10 mM HEPES pH 7.4 and 150 mM NaCl, resulting in three peaks. Pure protein was identified by SDS-PAGE in the third peak, which was concentrated to 7-10 mg/mL for further analysis. The overall yield of hLAMP3-prox was ~1.2 mg per litre cell culture for the RMCE cell clone (SWI3a-26a) and ~0.4 mg per litre for the FLEx cell clone (SWI4-25a). The higher hLAMP3-prox yield of the RMCE line indicated superiority of the RMCE expression locus over the FLEx locus.

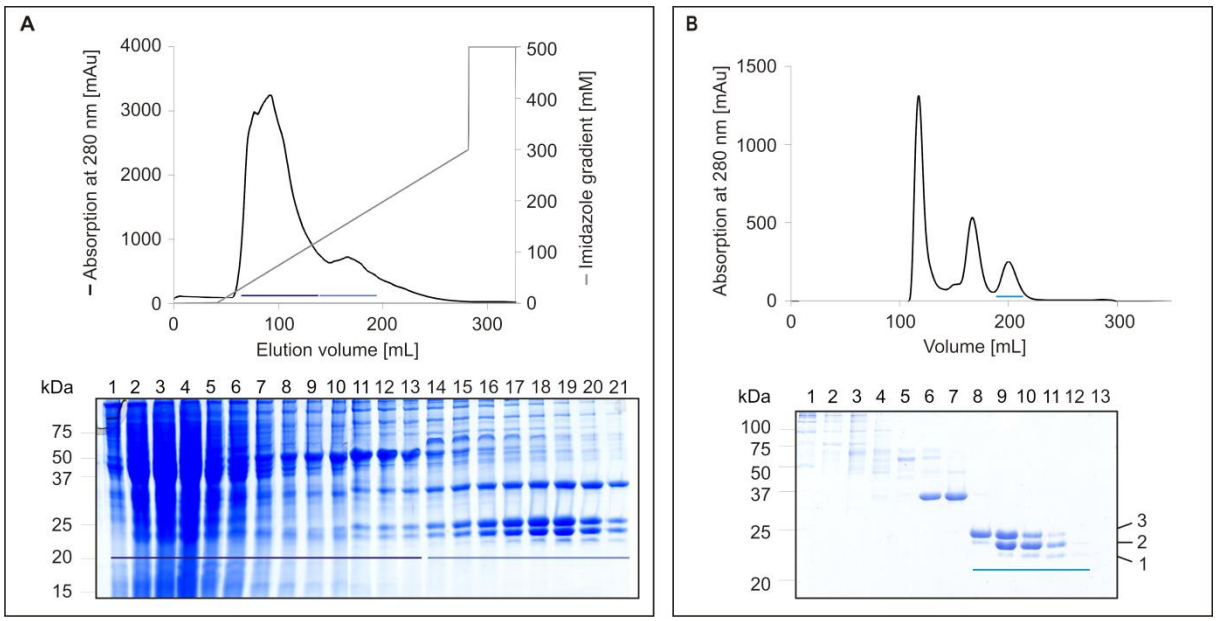


Figure 3-20: Purification of hLAMP3-prox. (A) hLAMP3-prox was produced in a bioreactor by a RMCE cell line. The perfused medium was concentrated tenfold and exchanged to phosphate buffer by diafiltration. The first purification step was an IMAC. Absorbance at 280 nm is indicated by a line in the chromatogram. Horizontal bars indicate fractions that were analyzed by SDS-PAGE. Fractions 14 to 21 contained hLAMP3-prox and were pooled (light blue bar). The main peak of the chromatogram (fractions 1 to 13, dark blue bar) contained unspecific bound protein from the culture supernatant. (B) Upon IMAC, hLAMP3-prox from pooled fractions was purified by GPC. Fractions 8 to 12 contained the pure protein. Proteins were analysed by 12% SDS-PAGE and Coomassie staining. 1 = hLAMP3-prox with one occupied N-glycosylation site; 2 = hLAMP3-prox with two occupied N-glycosylation sites; 3 = hLAMP3-prox with three occupied N-glycosylation sites.

Mass spectrometric peptide mapping of the three hLAMP3-prox SDS-PAGE bands between 20–25 kDa (Fig. 3-20 B) identified $GlcNAc_2Man_{4-5}$ at one (predominantly at site N291), two (predominantly at sites N266 and N291) or all of the three predicted N-glycosylation sites.

3.6.3 Deglycosylation of hLAMP3-prox

The deglycosylation of hLAMP3-prox was analysed by incubating the protein in the presence of different activities of endo H at pH 7.4 and pH 5.2, respectively, for different lengths of time. The digested protein was analysed by SDS-PAGE (Fig. 3-21). The lowest amount of endo H resulting in a complete deglycosylation was 2.5 U/µg when incubating over night at pH 5.2. However, aggregation of hLAMP3-prox in the acidic buffer condition was observed. hLAMP3-prox was therefore digested with 30 U/µg at pH 7.4 over night.

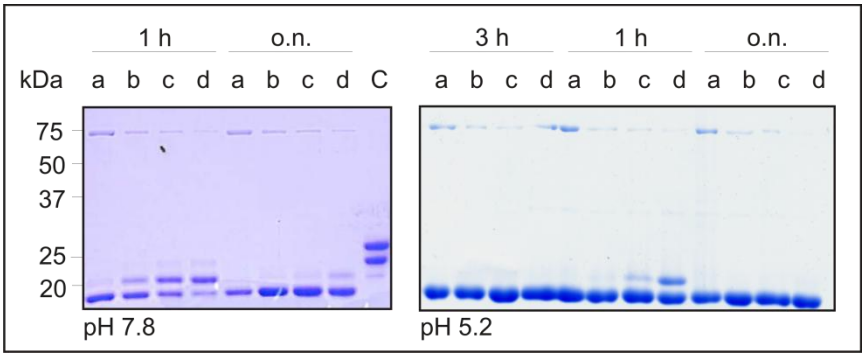


Figure 3-21: Optimization of hLAMP3-prox deglycosylation. hLAMP3-prox was incubated with different activities of endo H, for different lengths of time and at two pH values (7.8 and 5.2). a = 30 U/µg; b = 10 U/µg; c = 5 U/µg; d = 2.5 U/µg; C = hLAMP3-prox before deglysylation; o.n. = over night. Protein digestion was analysed by 12% SDS-PAGE and Coomassie staining.

Upon deglycosylation, hLAMP3-prox was purified by preparative GPC to remove endo H (Fig. 3-22). Endo H was identified in the fractions of the first peak, while the second peak consisted of pure, deglycosylated hLAMP3-prox, which was concentrated to 22 mg/mL.

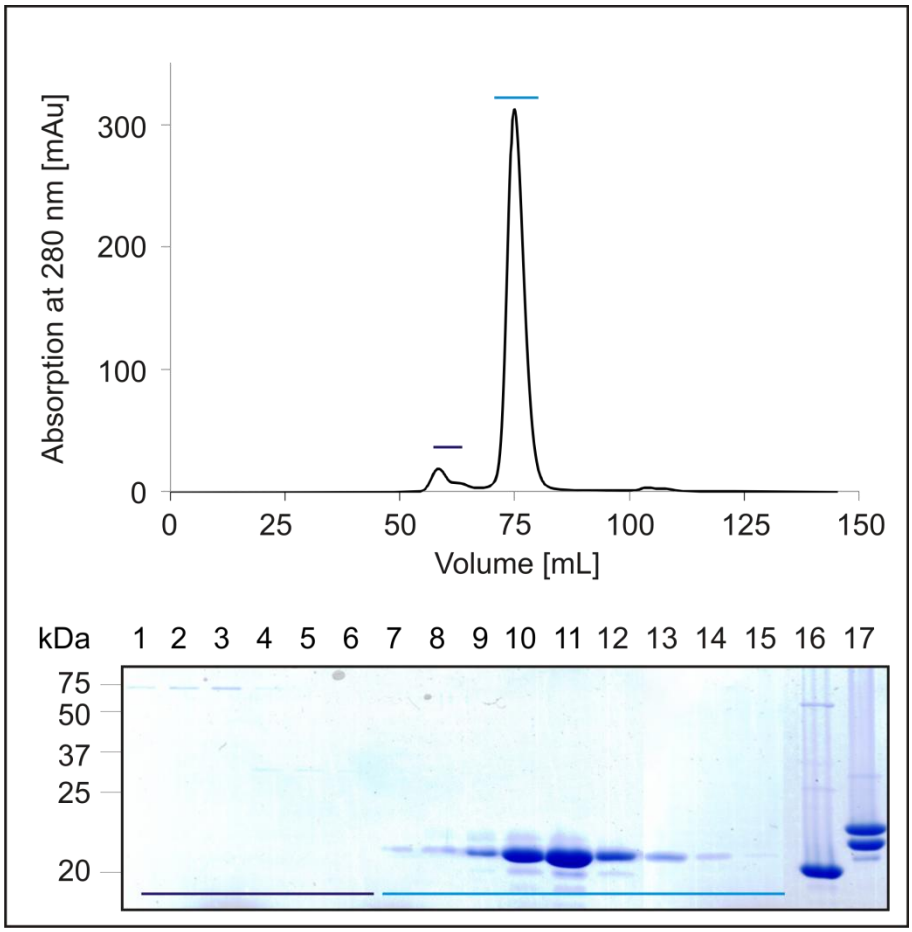


Figure 3-22: Purification of deglycosylated hLAMP3-prox. Upon deglycosylation, endo H was separated from deglycosylated hLAMP3-prox by GPC. Horizontal bars indicate fractions that were analyzed by SDS-PAGE. Fractions 1 to 6 contained endo H (dark blue bar); fractions 7 to 15 contained deglycosylated hLAMP3-prox (light blue bar). Lane 16 = deglycosylated hLAMP3-prox and endo H before chromatography; lane 17 = heterogeneously glycosylated hLAMP3-prox. Fractions containing protein were analysed by 12% SDS-PAGE and Coomassie staining.

Before crystallisation setups, the concentrated solution of purified and deglycosylated hLAMP3-prox in 10 mM HEPES pH 7.4 and 150 mM NaCl was analysed by dynamic light scattering (DLS) in order to check for aggregation and heterogeneity. Only one signal corresponding to a hydrophobic radius of 2.3 nm and a molecular weight of 23 kDa was obtained by DLS, which proved that the protein sample was monodisperse and free of aggregates (Fig. 3-23). However, it was observed that the deglycosylated protein was temperature sensitive that is to say it aggregated after a few minutes on ice or a few days at 4°C while it was soluble at RT for about two weeks. The glycosylated protein was stable at 4°C and RT over several weeks.

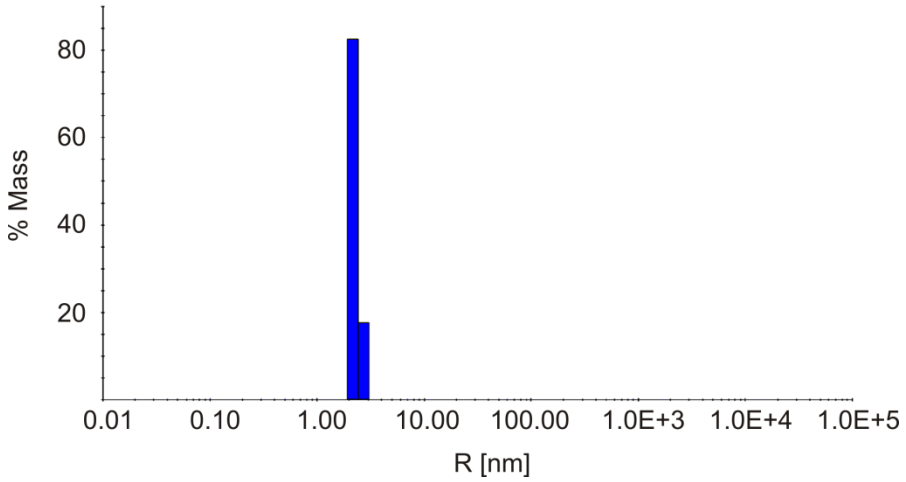


Figure 3-23: DLS measurement of purified and deglycosylated hLAMP3-prox. The presence of a single sharp peak around 2.3 nm indicated the absence of high-molecular-weight aggregates.

3.6.4 Crystallization of hLAMP3-prox

Initial screening of crystallization conditions for hLAMP3-prox was performed by sitting drop vapour diffusion using the following commercially available screens: JCSG core I, II, III and IV, the classics and the Pegs. 39 conditions yielded small, needle-like crystals after three to seven days at 19°C. Three representative conditions are shown in Figure 3-24 A to C. Crystals of 13 conditions were reproduced using an increased sample volume and hanging drop vapour diffusion in 24-well format (Table 3-5). This led to the crystals shown in Fig. 3-24, panels D-E, which diffracted X-rays to a maximal resolution of about 9 Å.

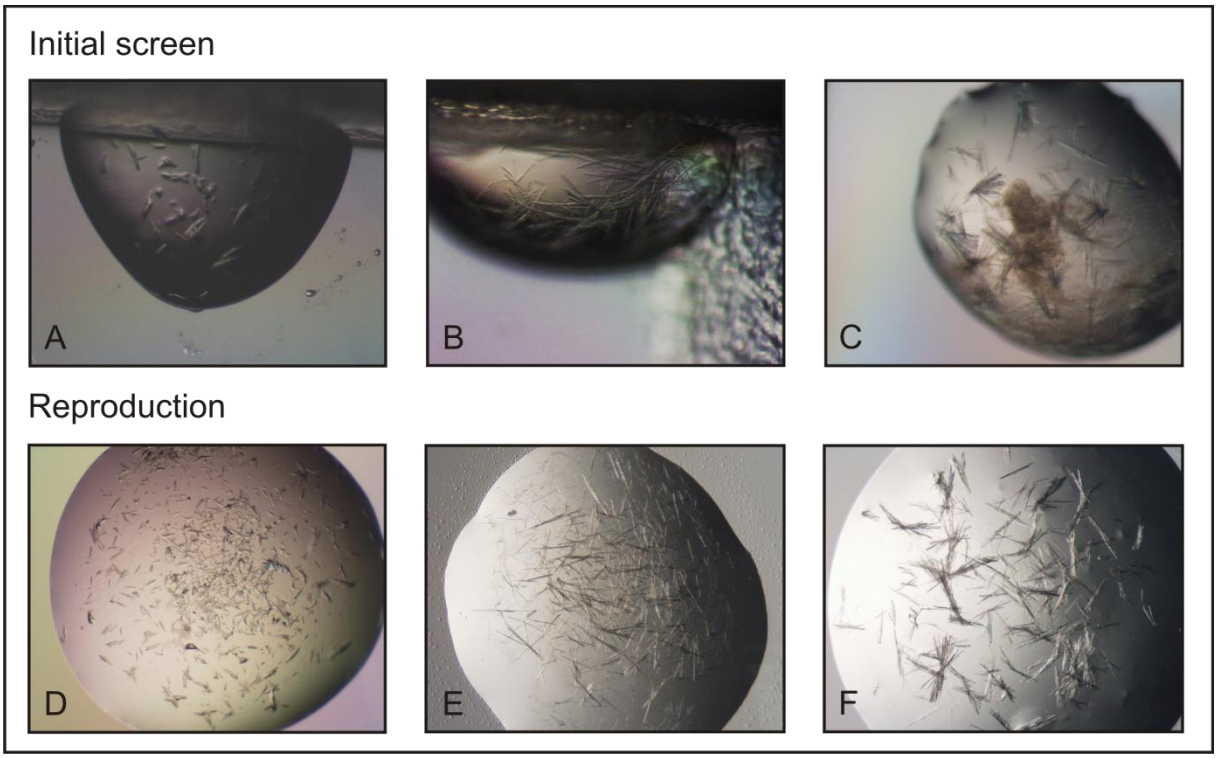


Figure 3-24: Initial hLAMP3-prox crystals. (A-C) hLAMP3-prox crystals were obtained from various crystallisation conditions using 96-well format crystallization screens. (D-F) Crystals from conditions presented in panels A to C were reproduced in 24-well format. Reservoirs contained (A, D) 0.2 M calcium acetate, 0.1 M MES, pH 6.0, 20% (w/v) PEG 8000; (B, E) 0.1 M citric acid, pH 5.0, 5% (w/v) PEG 6000; (C, F) 0.2 M ammonium sulphate, 20% (w/v) PEG 3350. Proteins were crystallized in drops of equal volumes of reservoir solutions and 22 mg/mL protein.

Table 3-5: Conditions that reproducibly yielded hLAMP3-prox crystals in 3 to 7 days.

No.	Conditions
1	0.1 M Na-acetate pH 4.6, 30% PEG2000, 0.2 M ammonium sulphate
2	0.1 M Na-acetate pH4.6, 25% PEG4000, 0.2 M ammonium sulphate
3	0.1 M Citric acid pH 2.5, 20%(w/v) PEG 6000, final pH 4.0
4	0.1 M Citric acid pH 5.0, 10% (w/v) PEG 6000
5	0.1 M Citric acid pH 5, 5% (w/v) PEG 6000
6	0.2 M Calcium acetate, 0.1 M MES pH 6, 20% (w/v) PEG 8000
7	0.2 M Ammonium sulphate, 20% (w/v) PEG 3350
8	0.1 M Sodium acetate anhydrous, pH 4.6, 20% PEG10000
9	0.1 M Sodium acetate anhydrous, pH 4.6, 15% PEG20000
10	0.2 M Magnesium chloride hexahydrate, 20% (w/v) PEG 3350
11	0.2 M Calcium chloride dihydrate, 20% (w/v) PEG 3350
12	0.2 M Ammonium sulphate, 20% (w/v) PEG 3350
13	0.2 M Ammonium dihydrogen phosphate, 20% (w/v) PEG3350

Optimization of the crystallization with 0.1 M citric acid pH 5, 5% (w/v) PEG 6000 was performed in 24 well plates in order to improve crystal quality and X-ray diffraction. The influence of varying the pH between 4.0 and 6.0, the concentration of the precipitant PEG 6000 from 3% (w/v) to 20% (w/v) and temperatures of 19°C and 26°C was tested. The first needle shaped crystals appeared again after three days and diffracted X-rays

up to 7 Å. Interestingly, in the following three months some crystals grew further while smaller ones dissolved according to a phenomenon called Oswald ripening (Ng *et al.*, 1996). Best diffracting crystals (up to 2.5 Å) appeared with the conditions 0.1 M citric acid pH 4.5 to pH 5 and 7% to 12% PEG 6000 at 19°C. Examples of various optimized crystals up to 750 µm in the longest dimension are shown in Figure 3-25. Streak or microseeding techniques in optimised conditions did not accelerate or improve the formation of the well diffracting crystals.

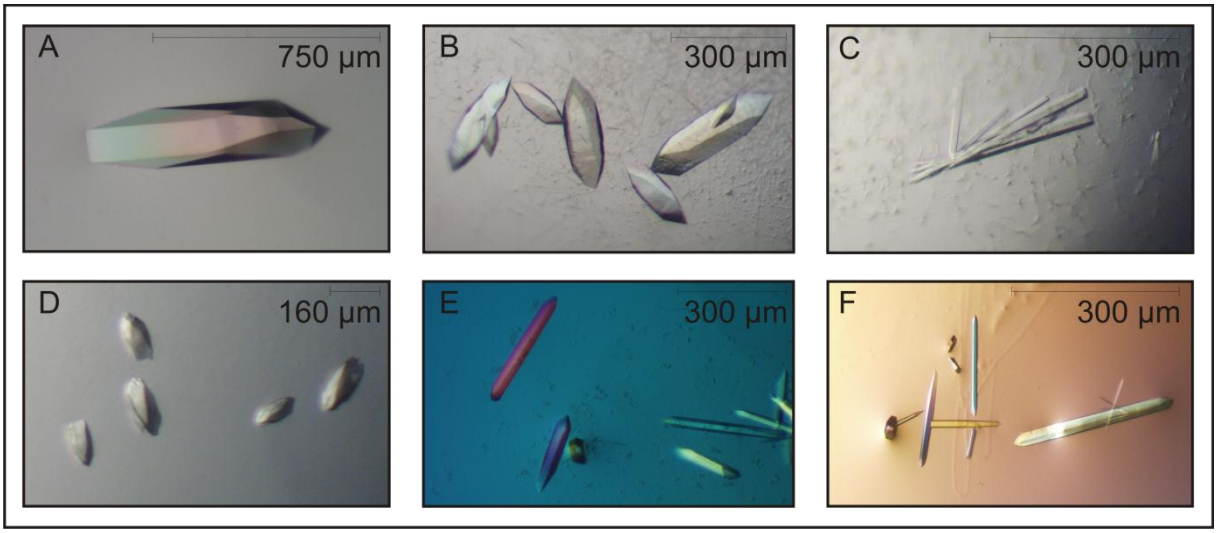


Figure 3-25: Optimized hLAMP3-prox crystals. Crystals were obtained through optimizing the screening condition 0.1 M citric acid pH 5, 5 % (w/v) PEG 6000 by varying pH and precipitant concentration. (A) 0.1 M citric acid pH 5, 7 % PEG 6000; (B) 0.1 M citric acid pH 4.5, 8 % PEG 6000; (C) 0.1 M citric acid pH 5, 14 % PEG 6000; (D) 0.1 M citric acid pH 4.9, 9 % PEG 6000; (E) 0.1 M citric acid pH 5, 12 % PEG 6000; (F) 0.1 M citric acid pH 5, 8 % PEG 6000.

Due to the long period of crystal growth, the intactness of the hLAMP3-prox construct building the crystal was analysed. To this end, crystals were dissolved and subjected to MALDI mass spectrometry (Fig. 3-26). The resulting molecular weight corresponded to the intact hLAMP3-prox construct bearing N-acetylglucosamine residues at two or three N-glycosylation sites. The glycoform with only one occupied site was not detected within the dissolved crystal solution.

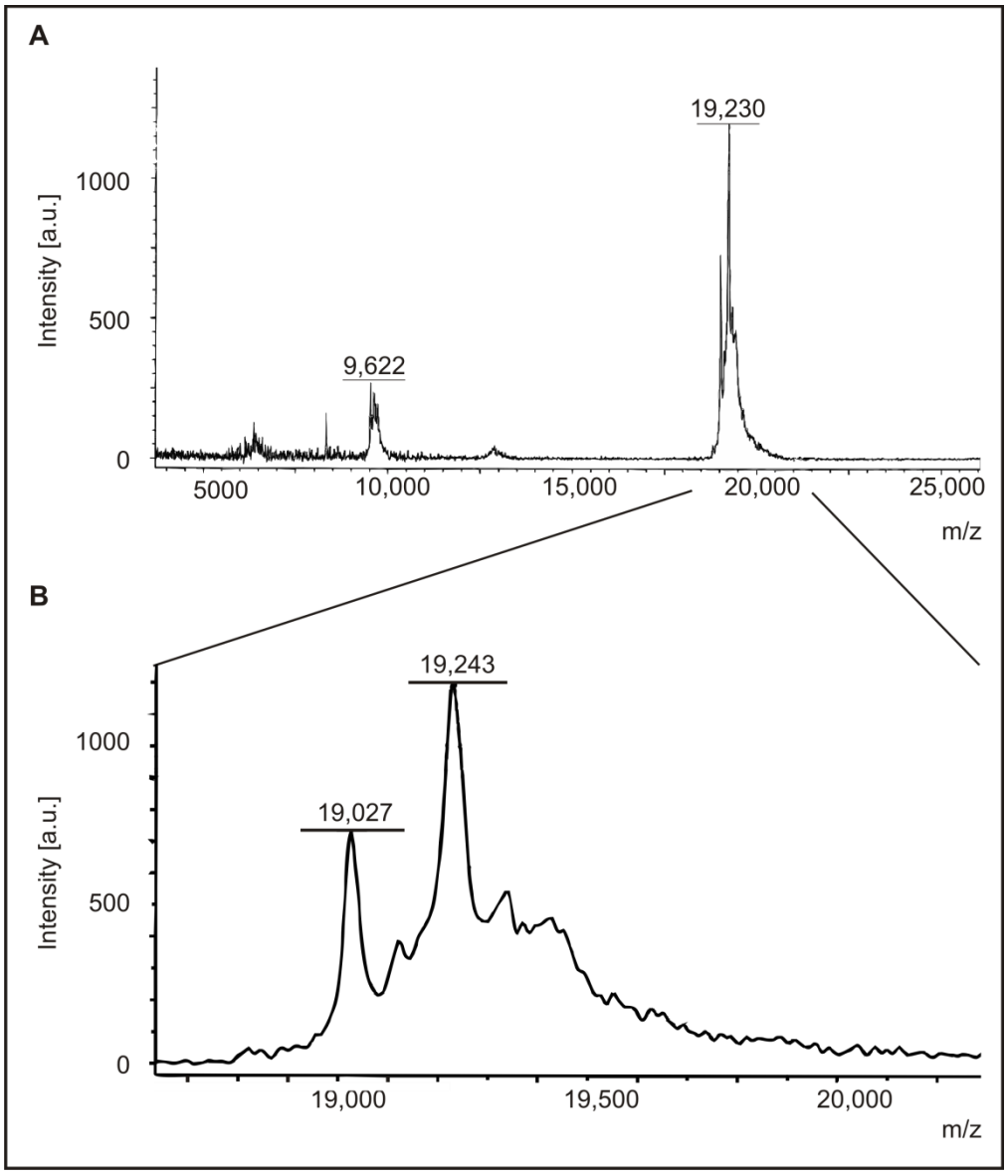


Figure 3-26: Characterization of hLAMP3-prox from a dissolved crystal by MALDI mass spectrometry. The molecular weight of 9,622 Da was assigned to a twofold positively charged standard molecule and 5000-7000 Da corresponds to PEG 6000 (upper panel) and molecular weights of 19,027 and 19,243, respectively, were measured for hLAMP3-prox modified by two GlcNAc residues and three GlcNAc residues, respectively (lower panel). m/z = ion mass/ ion charge; GlcNAc = N-acetylglucosamine.

3.6.5 Data collection and processing

Several optimized hLAMP3-prox crystals were used for X-ray diffraction data collection to determine the protein's three-dimensional structure. The number of recorded images and the oscillation range per dataset varied between different crystals. Both parameters were chosen based on the data collection strategy calculated by *iMOSFLM* (Leslie, 1992).

The crystals diffracted up to 2.7 Å at beamline X12 at the DESY (EMBL Outstation, Hamburg, Germany, Fig. 3-27). Indexing, integrating and scaling of the diffraction data

was performed by the programs XDS (Kabsch) and *XSCALE (Kabsch)*. Ice rings were excluded from the calculation. Trigonal symmetry (P3) with similar unit cell dimensions of a = b = 52 Å and c = 141 Å was determined for all isomorphous crystals. Moreover, the diffraction data indentified systematic absences along 00l. Only every third reflection was visible. The absences arose from destructive interference of X-rays in relation to a screw axis along c as it is present in the enantiomorphic space groups $P3_1$ or $P3_2$ (No. 144 or 145, $P3_x$) (Hahn, 1993). The complete data set statistics are shown in Table 3-6 at the end of this chapter.

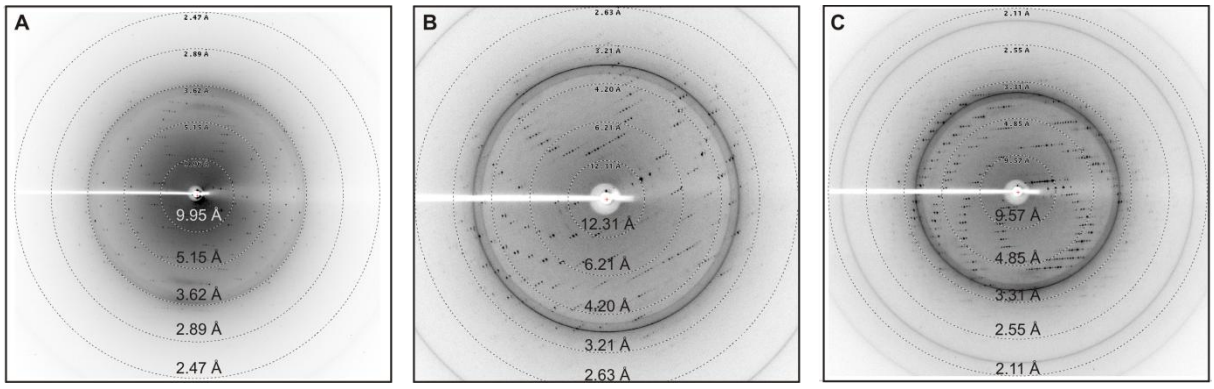


Figure 3-27: Diffraction images of hLAMP3-prox crystals. All diffraction images were collected at beamline X12 at DESY (EMBL Outstation, Hamburg, Germany). The resolution limits are indicated in the pictures. (A) Diffraction image of a sulphur SAD data set. (B) Diffraction image of a peak energy data set of an iridium derivative crystal. (C) Diffraction image of a peak energy data set of a platinum derivative crystal.

To estimate the number of protein molecules in the asymmetric unit, the Matthews coefficient was determined (Matthews, 1968). It indicated the presence of either two (V_M = 3.06 $Å^3$/ Da, solvent content: 59.87%, probability 36%) or, more likely, three (V_M = 2.04 $Å^3$/ Da, solvent content: 39.8%, probability 64%) monomers per asymmetric unit. The calculation of a MOLREP self-rotation function (CCPN4, 1994; Vagin & Teplyakov, 1997) revealed twofold axes in the xy-plane perpendicular to the crystallographic z-axis and a threefold axis parallel to the z-axis (Fig. 3-28). The threefold axis is identical to the crystallographic symmetry axis of space group $P3_x$, whereas the twofold axes correspond to a twofold non-crystallographic symmetry (NCS). This finding together with the Matthews coefficient indicates the presence of two hLAMP3-prox monomers as a C_2 dimer (Schönflies, 1886; Schönflies, 1887) in one asymmetric unit.

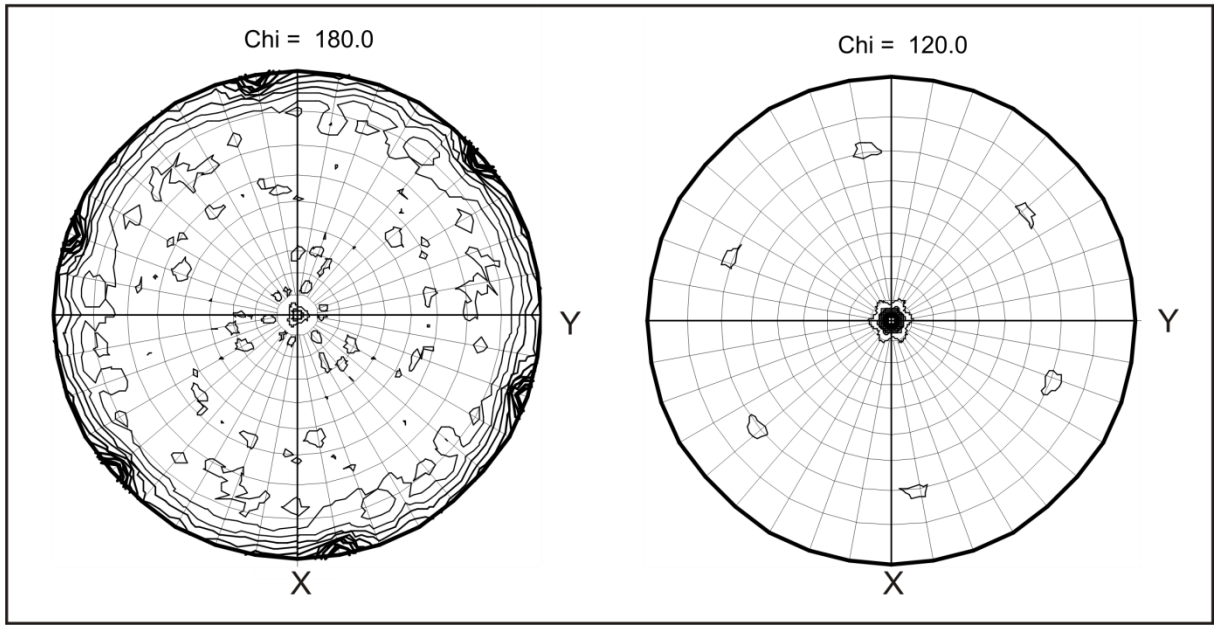


Figure 3-28: MOLREP self rotation function calculated from a data set of hLAMP3-prox. Stereographic projection of the coordinate surface of the spherical coordinates (r, ω, Φ) along the z-axis. Contour lines indicate the intersection point of a rotation axis. Left panel: search for twofold (χ= 180°) rotational symmetries, right panel: search for threefold (χ= 120°) rotational symmetries. One non-crystallographic twofold and one crystallographic threefold axis were found.

Upon scaling, corresponding data were converted with XDSCONV (Kabsch, 2010a; Kabsch, 2010b) to a MTZ file for further phasing analysis.

Table 3-6: Data collection statistics of LAMP-3 prox crystals. The values in parentheses account for the highest resolution shell.

Derivative	**Native**	**Iridium**			**Platinum**		
Data set	High-energy remote	Peak	Inflection	High-energy remote	Peak	Inflection	High-energy remote
Beamline	DESY X12	DESY X12	DESY X12	DESY X12	DESY X12	DESY X12	DESY X12
Wavelength (Å)	1.771000	1.103710	1.104200	1.100090	1.070450	1.070730	1.067870
Angular range (°)	360	180	180	180	180	180	180
Space group	$P3_1$ ($P3_2$)	$P3_1$ ($P3_2$)	$P3_1$ ($P3_2$)	$P3_1$ ($P3_2$)	$P3_1$ ($P3_2$)	$P3_1$ ($P3_2$)	$P3_1$ ($P3_2$)
Unit cell dimensions							
a (Å)	52.2	53	53	53	53	53	53
b (Å)	52.2	53	53	53	53	53	53
c (Å)	141.0	143.5	143.5	143.5	143.5	143.5	143.5
α (°)	90	90	90	90	90	90	90
β (°)	90	90	90	90	90	90	90
γ (°)	120	120	120	120	120	120	120
Resolution (Å)	20-2.7	20-2.8	20-2.8	20-2.8	20-2.6	20-2.6	20-2.6
	(2.86-2.7)	(2.95-2.8)	(2.95-2.8)	(2.95-2.8)	(2.72-2.6)	(2.72-2.6)	(2.72-2.6)
Unique reflections	22960 (3562)	19800 (3352)	19790 (3392)	19765 (3365)	26465 (3948)	26395 (4003)	13305 (2028)
Redundancy	5.76 (5.59)	2.84 (2.84)	2.83 (2.82)	2.84 (2.84)	2.89 (2.79)	2.83 (2.9)	5.74 (5.49)
Completeness (%)	96 (92.2)	90 (94.5)*	89.9 (96.9)*	89.7 (95.9)*	94.2 (88)*	94.5 (89.7)*	95.3 (91.1)*
I/σ_I	12.93 (2.71)	16.5 (3.45)	15.52 (2.92)	18.19 (3.84)	15.86 (2.0)	15.94 (2.0)	19.67 (2.2)
R_{merge}(%)	10.5 (58.6)	5.1 (30.1)	5.5 (39.5)	4.5 (27.8)	5.7 (63.8)	5.7 (64.8)	7.2 (83)
Mosaicity (°)	0.456	0.271	0.271	0.271	0.408	0.408	0.408

*Ice rings were excluded from data processing which affects the completeness of the data set.

3.6.6 Phasing

Since no appropriate model of LAMP-3 or the other LAMP family members was available, phasing by molecular replacement was not possible. Instead, anomalous dispersion was used to determine the phases of X-ray reflection experimentally.

Sulphur-single anomalous dispersion (S-SAD) phasing

First attempts to solve the phase problem by intrinsic sulphur SAD phasing failed due to macromolecular twinning (Fig. 3-29). The twinning fractions of the analysed crystals were measured to be 5% to 40% by SFCHECK (Vaguine *et al.*, 1999). Consequently, the anomalous signal-to-noise ratio (0.65) did not reach an arbitrary threshold of 1.3.

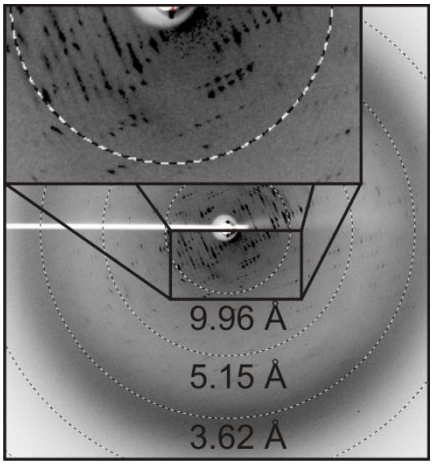


Figure 3-29: Diffraction image of a hLAMP3-prox crystal with a twinning fraction of 40%. The diffraction image was collected at beamline X12 at the DESY (EMBL Outstation, Hamburg, Germany). The resolution limits are indicated in the picture.

Multi-wavelength anomalous dispersion (MAD) phasing

Since phasing by S-SAD failed to solve the structure, multi-wavelength anomalous dispersion was used. Commonly, selenomethionine (SeMet) derivatized protein is produced and crystallized in such a situation. The SeMet derivatization in CHO Lec3.2.8.1, however, is impeded by the cytotoxicity of SeMet. Therefore, hLAMP3-prox crystals were derivatized by heavy-atom compounds. To screen for appropriate heavy-atom compounds for derivatization, native PAGE was performed by Vitali Maffenbeier which separates proteins on the basis of net charge, size and conformation (Boggon & Shapiro, 2000). The gel was loaded with soluble protein which had been incubated with different heavy-atom compounds. Two of these compounds potassium tetrachloroplatinate(II) (K_2PtCl_4) and ammonium hexachloriridate(III)-hydrate ($(NH_4)_3IrCl_6*xH_2O$) were identified to interact with hLAMP3-prox and thereby altering the protein's charge and thus the protein's electrophoretic behaviour. This is illustrated in Fig. 3-30 by the observed band shifts in lane two and four.

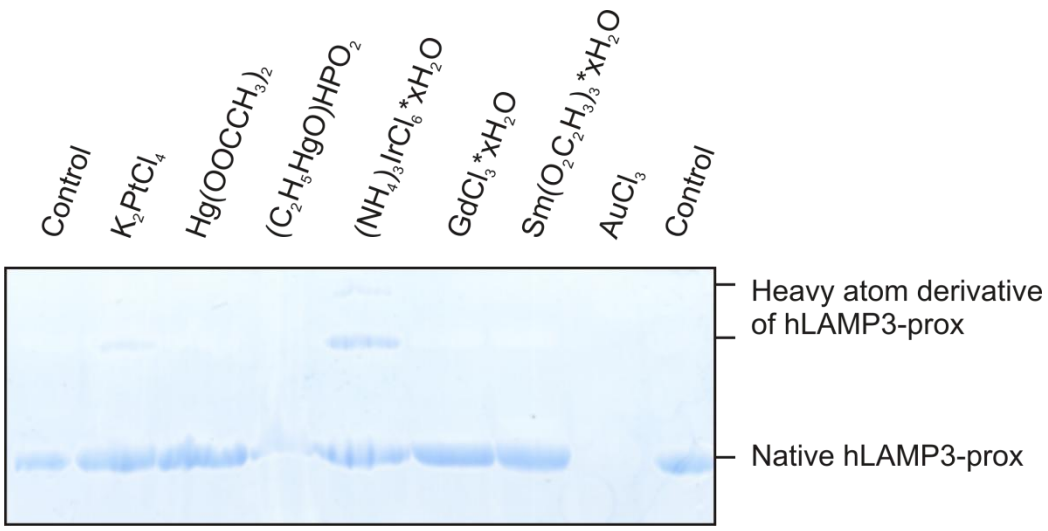


Figure 3-30: Native PAGE screening for heavy atom derivatization. 7 µg of the protein hLAMP3-prox were incubated at least one hour in 1 mM heavy atom solution in 0.1 M citric acid, pH 4.5. The reaction solution was analysed by a 15% native gel electrophoresis and coomassie staining.

The soaking of protein crystals with both components, however, easily provoked crystal cracking and loss of diffraction. To reduce the osmotic stress, back soaking in mother liquor for removal of the unbound heavy atom compounds was omitted. To determine the optimal wavelength for anomalous data collection, X-ray fluorescence energy scans of the soaked crystals near the Pt and Ir absorption edges was performed (Fig. 3-31 A). The resulting fluorescence data were used by the program CHOOCH (Evans & Pettifer, 2001) to calculate the anomalous scattering curve including the real (f') and imaginary (f'') part of the anomalous scattering factor (Fig. 3-31 B). To efficiently exploit the anomalous signal, three datasets were collected for each derivatized crystal at the wavelenghts marked in Fig. 3-31 A. Peak datasets were collected at the wavelengths of maximal f'', λ_{Ir}=1.103710 Å and λ_{Pt}=1.070450 Å for the Ir and Pt-derivatized crystals, respectively. A second dataset was collected at the inflection point (λ_{Ir}=1.104200 Å, λ_{Pt}=1.070730 Å), where the real part (f') of the signal reached its minimum (Fig. 3-31 B). Therefore, structure factors at this wavelength should strongly differ from structure factors at a wavelength which was chosen for the high-energy remote dataset (λ_{Ir}=1.100090 Å, λ_{Pt}=1.070730 Å). The gain of information from the latter two datasets alone (inflection/remote) is equivalent to two independent datasets collected from a native crystal and an isomorphous heavy metal derivative. The advantage of the MAD phasing is that all datasets were collected from the same crystal and therefore are perfectly isomorphous to each other. Non-isomorphism would have introduced systematic errors in the later phasing procedures.

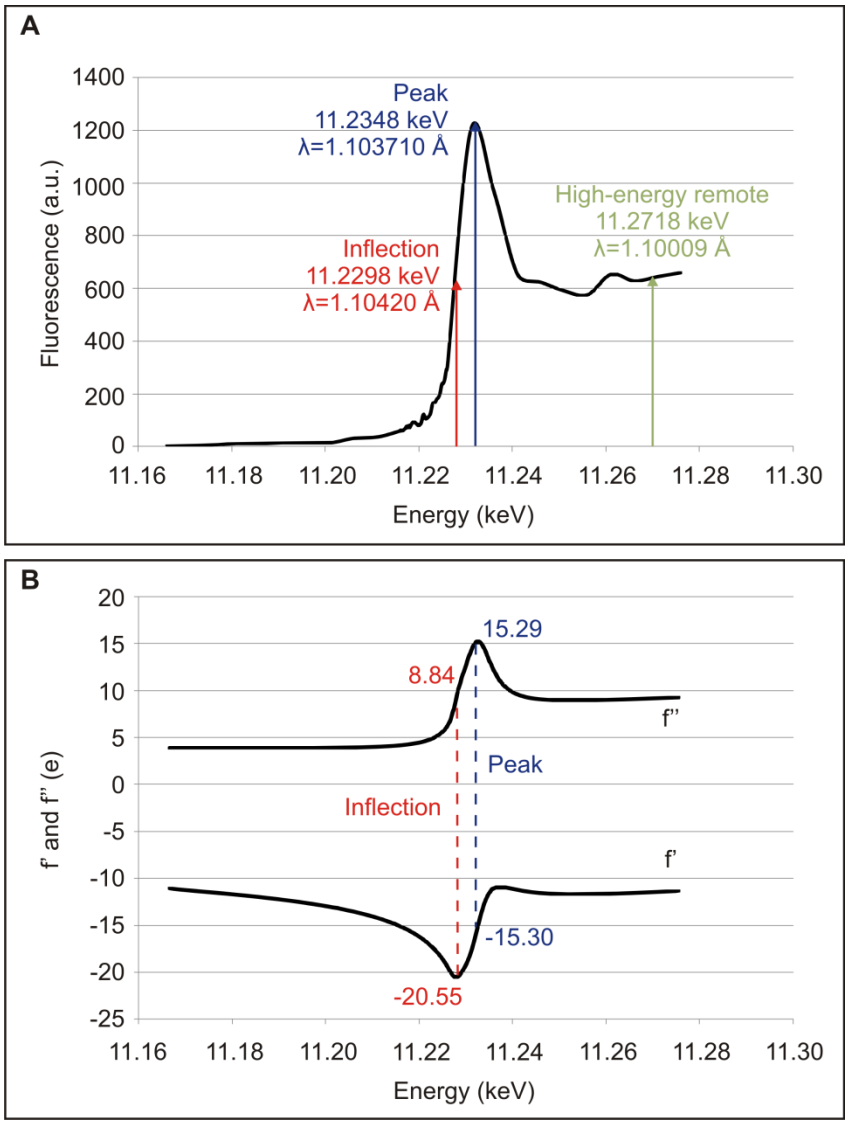


Figure 3-31: X-ray energy fluorescence scan of a hLAMP3-prox crystal derivatized with the iridium compound hexachloriridate(III)-hydrate. (A) The black line indicates the fluorescence intensity at a given energy of incident X-rays. Coloured vertical lines indicate the energies which were used for data collection. (B) Plot for the real (f') and imaginary (f'') part of the anomalous scattering factor at the L(III)-absorption edge of iridium. The values of f' and f'' at the inflection point and the peak are depicted in red and blue, respectively.

Following data collection and processing, experimental phasing calculations were performed by Dr. Jörn Krauße using the AutoSol tool of the PHENIX package (Adams *et al.*, 2010; Terwilliger *et al.*, 2009). Initial phases were eventually obtained by exploiting the anomalous signal of the iridium data set up to a resolution of 2.8 Å where the anomalous signal-to-noise ratio reached the threshold 1.3. In contrast, the platinum data set could not successfully be used for phasing due to a too small anomalous signal. Thus, the initial model was obtained by the iridium data set having incorporated one iridium atom per molecule (Fig. 3-32). It is coordinated with the nitrogen of S346

(distance 4.1 Å), the carbonyl oxygen atoms of Q320 (distance 3.9 Å), A347 (distance 4.0 Å), H348 (distance 3.6 Å) and probably the oxygen atoms of two water molecules.

Initial hLAMP3-prox structure

In the presented model hLAMP3-prox appears as a monomer. Most of its amino acids were built with the exception of a loop region comprising seven amino acids (V230 to L236 referring to their position in the full length protein) near the protein's N-terminus. Here, the electron density was poorly defined which might be improved by further refinement and new crystals with higher resolution.

The model in Figure 3-32 is in agreement with the secondary structure prediction (Fig. 3-18) and reveals a high β-strand content. Anti-parallel β-strands form a sandwich of two β-sheets. Hydrophobic interactions between the β-sheets (Fig. 3-32 B) and two disulphide bonds (C237-C274 and C339-C376, Fig. 3-32 A) stabilize the domain. Moreover, electron density corresponding to GlcNAc residues was measured at glycosylation sites N266 and N291 whereby the electron density around N266 was not well defined. This can be due to heterogeneous occupation of this glycosylation site or the flexible nature of the carbohydrate residue. The GlcNAc bound to amino acid residue N291 forms a crystal contact over its C3-hydroxyl group to the nitrogen of a crystal neighboured G287 backbone and vice versa (Fig. 3-33). Thus, this hydrogen bond stabilized the crystal. The third glycosylation site at N232 is localized in the loop that had not been built yet.

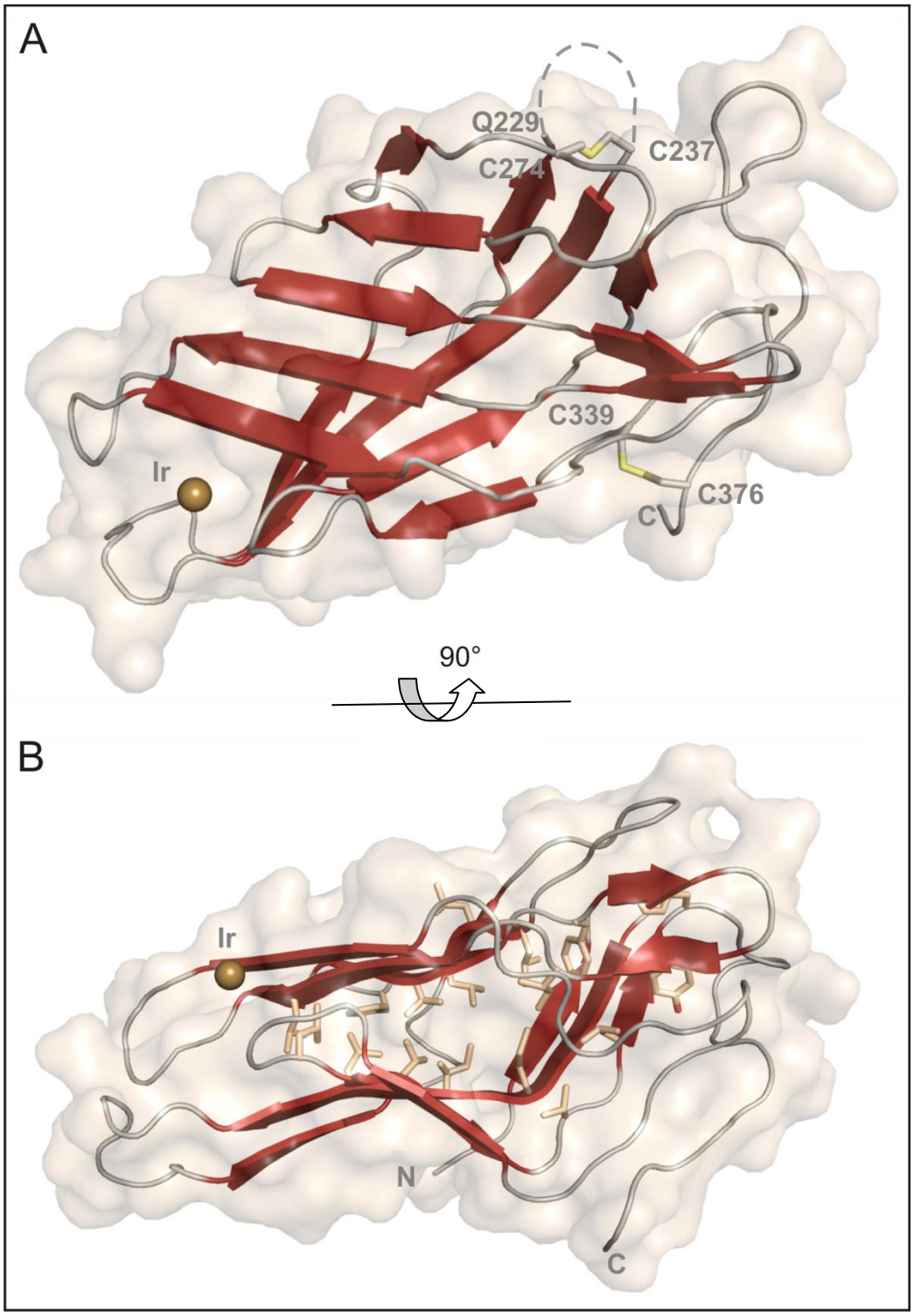


Figure 3-32: Initial model of hLAMP3-prox. Cartoon representation of the overall structure. Two views are shown related by a 90° rotation around the horizontal-axis. (A) Positions of cysteines 237, 274, 339 and 376 involved in disulphide bonds are marked and the sulphurs are indicated in yellow. (B) Hydrophobic amino acid side-chains inside the sandwich fold are shown in stick representation. Loop regions are shown in grey and β-strands in red. Iridium atoms are represented by brown spheres. Figures were generated with Pymol (http://pymol.org).

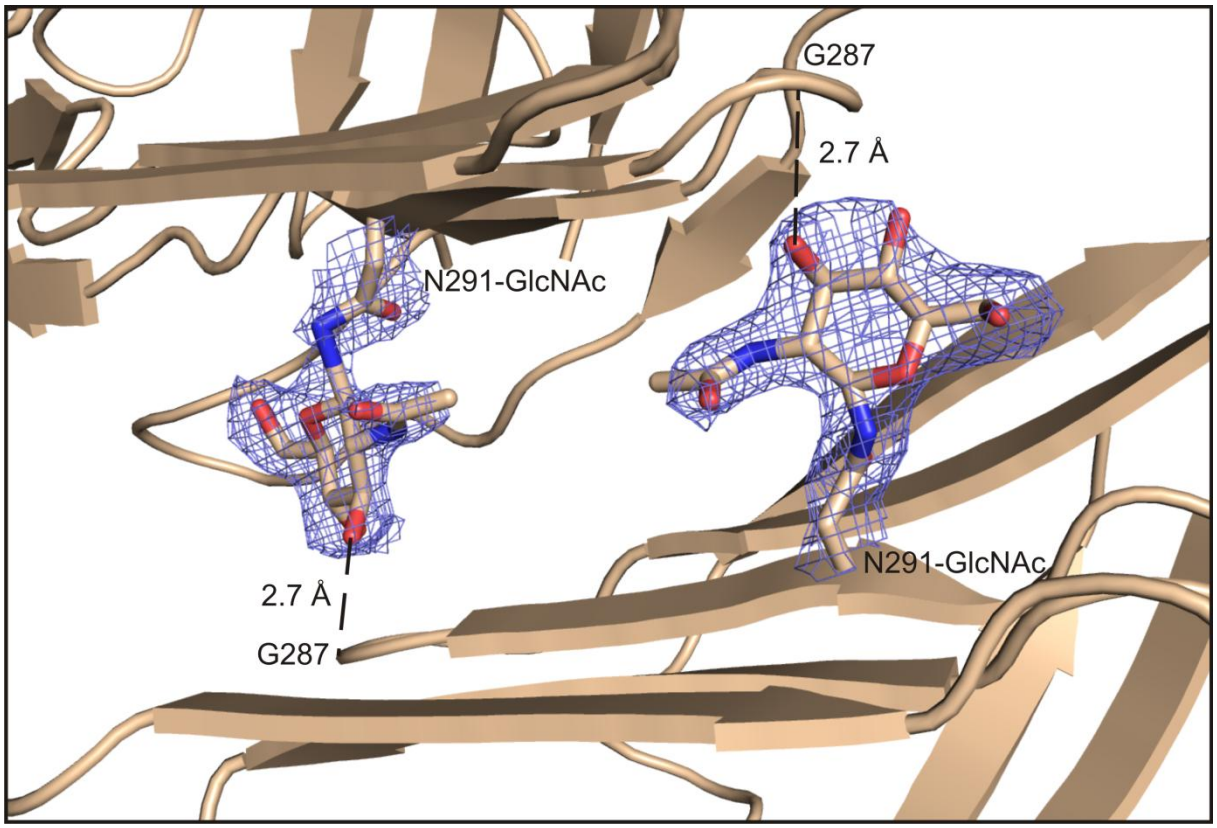


Figure 3-33: Crystal contact in hLAMP3-prox crystals. A hydrogen bond (2.7 Å) is formed between the C3-hydroxyl group of the GlcNAc bound to residue N291 and the nitrogen of the G287 residue located in a crystal neighboured molecule. The figure was generated with Pymol (http://pymol.org).

To gain further information about hLAMP3-prox, the presented model was used to search for structurally related proteins in the protein data bank using the DALI server (Holm & Sander, 1999). Significant similarities have a *Z*-score above 2 and usually correspond to similar folds. Strong matches even have a *Z*-score above a cut-off that is empirically set to *n*/10 – 4, where *n* is the number of residues in the query structure (Holm *et al.*, 2008). The protease VII of *E. coli* had the highest Z-score of 5.2 but no obvious similar fold. Thus, no convincing structural homologue was found. Further refinements leading to an improved electron density and a more complete model might improve the DALI search. Beyond that, the presented model is the basis for further experiments resulting in a more detailed structural insight into this protein and other members of the LAMP family.

4 Discussion

4.1 Site-specific recombination systems in cell line development for structural studies

A variety of approaches have been evolved to enhance recombinant protein production in mammalian cells. However, stable cell line development is still restrained by time-consuming and expensive screening procedures for high and stable production clones. Defined genomic engineering by heterologous site-specific recombination systems has been described before (Kaufman *et al.*, 2008; Nehlsen *et al.*, 2009; Qiao *et al.*, 2009; Schlake & Bode, 1994). Published reports have so far focused on the proof of principle and optimization of the technology. This work describes for the first time the application of recombination systems to establish glycosylation mutant cell lines for protein production for crystallography. To this end, glycosylation mutant CHO Lec3.2.8.1 cells were chosen as host cells. These cells express glycoproteins with truncated and more homogenous carbohydrate side chains which have a beneficial effect on crystallisation. Two different recombination strategies were pursued: an Flp-mediated excision (FLEx) and a recombinase-mediated cassette exchange (RMCE). In both approaches, chromosomal loci were tagged by Flp recombination target (FRT) sites and a GFP gene. The GFP gene was used to enrich and clone highly expressing cell clones using fluorescence activated cell sorting (FACS). Upon generation of so-called master cell clones, the GFP reporter was either exchanged against another gene (RMCE) or excised (FLEx) whereby a gene of interest was put under control of the promoter at the pre-characterized locus. By this means, production cell lines were established for different recombinant products, and they were successfully used for bioreactor cultivations.

In the following the results of the RMCE and the FLEx system are discussed and compared regarding their applicability for glycoprotein production in crystallography.

4.1.1 Isolation of genomic loci with stable and high transgene expression

An important characteristic of a good production cell line is its highly active and stable expression locus. Conventional identification of an expression hot spot is based on random integration of the transgene resulting in unpredictable integration sites and transgene copy numbers due to chromosomal position effects. This explains the highly

variable gene expression among isolated cell clones, requiring an extensive screening procedure. Moreover, the selection procedure has to be repeated for every new recombinant product which in turn impedes a higher throughput. Hence, new strategies are of great interest.

Various strategies have been applied before to identify expression hot spots. They include antibiotic selection, multiple rounds of FACS and combinations of both (Kim & Lee, 2008; Nehlsen *et al.*, 2009; Qiao *et al.*, 2009; Verhoeyen *et al.*, 2001). FACS enables high-throughput selection of high-yield GFP tagged clones out of millions of cells and facilitates an easy long-term screening of cell clones for homogenous expression (Kaufman *et al.*, 2008). Thus, this method was well suited for establishing master cell lines in this work. After nucleofection of CHO Lec3.2.8.1 cells a two-round sorting strategy resulted in stable, high-yield expression in 38% of the analysed clones after 12 weeks in culture and in the absence of any antibiotics (Fig. 3-3 B). The observed reduction in productivity in some clones may be a consequence of chromosomal rearrangements (Yoshikawa *et al.*, 2000) or epigenetic mechanisms occurring by interactions between the integrated DNA sequence and the chromosomal surroundings (Emerman & Temin, 1984; Whitelaw *et al.*, 2001). Besides an emphasis on position effect, tandem-repeat integration of the transgene may lead to gene silencing (McBurney *et al.*, 2002). The transgene copy number influences the cassette exchange, too. While “head-to-tail” tandem integrations result in the excision of additional transgenes by recombination of identical FRT sites, more complicated architectures may result in unpredictable deletion, inversion or translocation (Bode *et al.*, 2000). To circumvent this unpredictability, transgene copy numbers and tandem-repeat integrations of RMCE master cell clones were investigated by PCR and Southern Blot analysis. Among the analysed 34 cell clones, 8 (26 %) and 3 (20 %) were identified with a single integration of pEF-FS-EGFP and pEF-FS-GFP-dneo, respectively (Fig. 3-8 D). This result suggests that nucleofection is a feasible DNA transduction method for CHO Lec3.2.8.1 cells in order to obtain single copy integrations. Similar single copy integration rates of about 30 % were obtained for CHO K1 cells by Nehlsen *et al.* (2009).

With regard to the cell sorting strategy itself, it is often recommended to use a destabilized, short-lived GFP or RFP variant (Turan *et al.*, 2011). This is advantageous,

if the enhanced GFP variant reaches saturation and prevents resolution of the top fluorescent cells. Figure 3-2 B shows, however, that only a minor part of the GFP cells reached saturation in this work. The maximal GFP expression levels ranged between 11 mg/L and 30 mg/L. These cells were considered to be adequate master cells for glycoprotein production in structural biology. Altogether, 12 and 18 FLEx master cell lines were generated for the target proteins scHGF and hLAMP3-prox. 8 and 3 single copy master cell lines were generated for the RMCE strategies based on the plasmids pEF-FS-EGFP and pEF-FS-EGFP-dneo, respectively.

4.1.2 How to efficiently isolate recombination events upon RMCE

The cassette exchange that was applied in this work based on the Flp/FRT system of *Saccharomyces cerevisiae*. Here, a gene cassette is flanked by two FRT sites, a wild-type and a mutant sequence, that cannot recombine with each other but are recognized by Flp (Schlake & Bode, 1994). This architecture enables the precise exchange of one gene cassette against another. However, the exchange rate is low. The simplest selection approach aimed at sorting the exchanged events by FACS. Since the GFP reporter is exchanged against another gene, the cells should lose GFP fluorescence and can be isolated. Four master cell clones with different integration loci were investigated for an exchange of GFP against RFP. However, no RFP^+, GFP^- cells were obtained. This regative result led to the introduction of antibiotic selection strategies.

Exchanging GFP against a hygromycin resistance thymidine kinase fusion gene (HygTk) in five different master cell clones resulted in two to 600 mostly GFP^+, but also GFP^- colonies after hygromycin B (hyg B) selection. Though being leaky, the hygromycin selection system was adapted to generate scHGF production clones. The exchange vector bore the scHGF gene followed by an IRES element and the HygTk gene for selection. However, this approach failed in growing GFP^- colonies. The high number of selected GFP^+ colonies indicated that the master cells became resistant against hyg B on another way than RMCE. The exchange vector might have been randomly integrated downstream of an endogenous promoter and thereby conferring resistance. Due to the high frequency of this event interfering with RMCE, a more stringent selection system was needed. A combination of a positive and negative selection strategy was for example conceivable. One possibility would have been the use of generated HygTk clones as master cells. An exchange of HygTk by a GOI could be negatively selected by ganciclovir, which is a pro-drug excluding non-exchanged

parental cells from growing due to the viral thymidine kinase expression (Seibler *et al.*, 1998). The tymidine kinase converts ganciclovir into a phosphorylated nucleotide analog, which incorporates into the DNA of replicating eukaryotic cells and causes their death (Tomicic *et al.*, 2001). A published approach in BHK-A cells made use of a combined positive and negative selection strategy by puromymicin and ganciclovir, respectively (Karreman *et al.*, 1996). In that work, 62 cell clones were established for two integration sites, but in many cases they bore random integrations of the exchange vector, too. Only one was identified as an authentic RMCE event. Thus, the negative selection strategy was not followed further in this work.

Higher recombination efficiencies were obtained by integrating a selection trap in form of a non-functional selection marker like the neomycin resistance gene (Δneo^R) (Verhoeyen *et al.*, 2001). The Δneo^R gene is cloned directly downstream of the FRT cassette in the master cells and lacks a promoter and a start codon. Upon recombination, the missing elements, located on the targeting vector, transcriptionally activate the incomplete Δneo^R gene at the tagged locus (Fig. 1-9). The complemented selection trap allows selection for correct RMCE events with the antibiotic G418. Verhoeyen (Verhoeyen, 2000) analysed 165 subcell clones upon RMCE and confirmed a correct exchange in all of them. Schucht (Schucht, 2006) used the Δneo^R selection trap in three cell lines (HEK293, BHK-21 and CHO-K1) and also found nearly 100% exchange efficiency. Only three of 122 analysed subclones contained additional integrations of the exchange vector. Thus, this selection system facilitated a stringent selection of authentic cassette exchanges.

This work combined the Δneo^R complementation strategy with a promoter trap (Fig. 1-9). No targeting vector integrated randomly in 15 of 15 analysed non-fluorescent subcell clones. Furthermore, RMCE resulted in a complete exchange in 43 of 45 subclones derived from two different single-copy master cell clones (SWI3a-26 and SWI3b-5). Thus, the cassette exchange in CHO Lec3.2.8.1 cells was regarded as reliable and obviously resulted in extremely high selection stringency.

4.1.3 RMCE frequency

Besides selection stringency, recombination frequency contributes to an efficient RMCE experiment. The latter is influenced by the recombinase activity, the exchange vector and the position effect of tagged chromosomal loci in the master cells.

As the activity of the natural thermo-instable Flp was low in mammalian cells, the thermo-stable mutant Flpe was evolved by cycling mutagenesis (Buchholz *et al.*, 1996a). The recombinase that was used throughout this work based on a mouse codon optimized Flpe variant (FlpO) whose efficiency is comparable to Cre (Raymond & Soriano, 2007). Recently, it was reported that the exchange efficiency is directly related to the amount of transferred Flp expression plasmid (Schebelle *et al.*). The following amounts of Flp expression vector per $5x10^6$ cells were transferred in other studies: 170 µg for murine ES cells (Seibler *et al.*, 1998), 40 µg for CHO cells (Qiao *et al.*, 2009) and 25 µg for NIH3T3 cells (Turan *et al.*, 2010). In this work, however, the transferred DNA amounts were limited to 5 µg and a maximum volume of 5 µl by the DNA delivery method nucleofection. The corresponding recombination frequencies varied between 7 and 50 subcell colonies per $1.5x10^6$ transfected cells dependent on the used exchange vector. Best results were obtained by transfecting the Flp expression vector and the exchange vector in a molar ratio approaching 2:1. The excess of the Flp enzyme is necessary regarding the stoichiometry of the recombination reaction. Four Flp protomers have to associate with the DNA strands of both reaction partners for catalysing the recombination (Chen & Rice, 2003). Moreover, the protein dissociation step limits the turnover of Flp reactions (Waite & Cox, 1995). Thus, Flp remains unavailable for further reactions for extended time periods and higher amounts of the enzyme can compensate for the deficiency.

A prerequisite for an Flp-mediated recombination is a random collision of two identical FRT sites (Qiao, 2009). In case of RMCE, one FRT cassette is anchored as a single-copy in the host chromatin, while its counterpart is located on the exchange vector. Thus, the amount of the incoming FRT cassette on the exchange vector can be varied. By increasing its amount to a large molecular excess, the rate of random collisions possibly leading to RMCE events is raised. Higher recombination efficiencies may be achieved by optimizing other DNA delivery methods for the hard-to-transfect CHO Lec3.2.8.1 cells. DNA transfer by polyethyleneimine (PEI) could be a solution as it is

cheap and the DNA amounts can be increased. The ratio of DNA to PEI for CHO Lec3.2.8.1 cells was optimized in small scale and adequate transfection rates of up to 29% were obtained with 1.6 µg of an GFP reporter plasmid and 20 µg PEI per $5x10^5$ cells (Duda, 2009). However, it has to be taken into account that a transfer of increased DNA amounts might also raise the probability of random secondary integrations of the Flp and the exchange vector.

The most obvious difference between RMCE runs was the transgene size and therefore the size of the exchange vector. While transfection with an exchange vector bearing a RFP gene (0.68 kb) resulted in 40-50 colonies, transfection with an exchange vector bearing a scHGF gene (2.2 kb) resulted in 7 colonies. This led to the assumption that the distance between two FRT sites might influence the rate of collisions between the respective FRT sites. For Flp-mediated excision approaches, where both FRT sites were located on one DNA molecule, it was shown that increased distances between the FRT sites led to an asymptotic decrease in recombination frequency (Ringrose *et al.*, 1999). Nevertheless, Wallace *et al.* (Wallace *et al.*, 2007) described the replacement of a large segment (> 100 kb) of the mouse genome by the equivalent human synthetic region. This shows RMCE being a multi-factorial process that can be optimized at different levels to enable reproducible results.

No significant variation was observed when master cell clones with different integration loci but similar high GFP fluorescence were transfected by the same RFP exchange vector. The recombination frequency was uniformly 40 to 50 colonies per approach. This supports the observations of Verhoeyen (Verhoeyen, 2000) and Schucht (Schucht, 2006) that RMCE frequency is similar for different high expression integration sites. Highly producing cells were more amenable for site-specific recombination than lower producing cells in their studies. It appears that tagged chromosomal loci located in easily accessible heterochromatin conferring a high transcriptional activity also support success of RMCE.

4.1.4 RMCE of multiple transgenes is possible

Diverse studies have shown before that the co-expression and co-purification of protein complexes for structural studies are often more successful than their separate handling

(Romier *et al.*, 2006; Strong *et al.*, 2006). Thus, a system facilitating the co-expression of multi protein complexes in mammalian cells would be of great interest.

In this work, RMCE was performed with five multi-copy master cell lines (Table 3-1) and resulted in a complete exchange of all reporter copies against RFP in three lines. This suggests that simultaneous exchange of two to three reporter gene copies against different target genes for co-expression in the same cell would generally be possible. However, the selection of subcells expressing all desired transgenes might be challenging using the master cells of this work, as all loci bear the Δneo^R trap. The integration of different selection traps such as a hygromycin trap or puromycin trap might improve selection specificity. Further considerations have to be made before establishing a robust RMCE system for the expression of multi-protein complexes at different chromosomal loci (multiplex RMCE). For example, mutual recombinations between the FRT sites of different tagged loci have to be ruled out. Recently, a multiplex Flp-RMCE approach was described that relies on a set of novel heterospecific FRT site mutants (Turan *et al.*, 2010). The authors designed novel pairs of synthetic FRT site variants and evaluated simultaneous exchange of two separate target cassettes in one master cell. The new F13/F14 site pair was similarly effective in RMCE as the F3/F sites used in this work.

Besides the three multi-copy master cell clones that exchanged all FRT-cassettes, two multi-copy clones did not. False subclones were usually identified by their GFP fluorescence. However, one master cell line, SWI3a-33, always resulted in subcell clones with incomplete RMCE events, though their subcell clones lacked green fluorescence and produced glycoproteins with good yield. This may be explained by a transcriptionally silenced copy of the reporter gene present in the heterochromatin, not accessible for recombination. In summary, unpredictable RMCE results can be expected when using multi-copy master cells with identical FRT-cassettes.

4.1.5 Homogenous expression by isogenic subcell clones

According to the RMCE principle, isogenic subcell clones of one master cell line should have a uniform expression profile after cassette exchange. Previous studies showed this for the production of retroviral vectors and antibodies in HEK293 and CHO cells (Nehlsen *et al.*, 2009; Schucht *et al.*, 2006).

The genomic homogeneity that was found in subclones of master cells with single copy transgenes corresponded to homogenous RFP levels (Fig. 3-12). Interestingly, the RFP levels are quite similar throughout all analysed subclones, though the GFP levels of different master clones varied. Regarding the position effect, there are no differences between master cells and their daughters except for the transgene itself (Fig. 1-9). Thus, there is the possibility that particular features of the transgene like its nucleotide composition or its product might limit the RFP concentration in mammalian cells. For bacterial hosts it is for example discussed that the RFP variant DsRed-Express, which was also used in this work, is cytotoxic due to aggregation (Strack *et al.*, 2008).

Close analysis of five RMCE-derived scHGF cell clones revealed deviations in their production level determined by ELISA, although immunoblotting resulted in similar scHGF expression. These discrepancies coincided with differences in cell size, growth and DNA content. Non-uniformity of scHGF cell clones was likely caused by chromosomal abnormalities like aneuploidies of the CHO genome, which is known to be instable (Barnes *et al.*, 2003; Pilbrough *et al.*, 2009). All other cell lines obtained by RMCE in this study, apart from scHGF, appeared to be isogenic and interchangeable on protein level. Thus, CHO Lec3.2.8.1 hosts were generally well amenable for the Flp RMCE strategy. Nevertheless, the size of subclones obtained by RMCE should be taken into account. Subclones of similar size to the master cells should be selected to ensure predictable production and cell growth.

In the following, another aspect leading to heterogeneity of subclones in other heterologous recombination systems will be discussed. The Cre/lox system from bacteriophage P1 has been reported to cause decreased cell growth, cytopathic effects and chromosomal aberrations (Schmidt *et al.*, 2000; Silver & Livingston, 2001). Explanations for these observations were illegitimate recombinations based on the presence of multiple cryptic lox sites on the one hand. On the other hand illegitimate endonuclease activity contributed to DNA damage and growth inhibition by inducing nicks that can be converted to double strand breaks and led to non-homologous DNA repair (Loonstra *et al.*, 2001). The illegitimate endonuclease activity can arise, as recombination intermediates are prone to be attacked by unauthorized nucleophiles like water molecules. In case of the Flp/FRT system this side reaction is confined by the architecture of the recombination synapse and its conformational dynamics as identified

by three dimensional structures of Flp (Ma *et al.*, 2009). Furthermore, a long term study proved the stability of an FRT flanked gene cassette in a BHK cell line continuously expressing Flp (Seibler & Bode, 1997). This strongly indicated the absence of cryptic FRT sites in mammalian genomes. Taken together, the Flp/FRT system confers highly specific and non-toxic recombination (Turan *et al.*, 2011). Thus, the heterogeneous scHGF expression levels found in this work are more likely to be caused by characteristics of the CHO Lec3.2.8.1 cells.

4.1.6 Flp-mediated cassette exchange versus excision

In this work, stable high producer cell lines for two target proteins, hLAMP3-prox and scHGF, were successfully established by RMCE and FLEx. Their establishment relied on two steps. The first step was the same for both systems: integration loci conferring high and stable GFP reporter gene expression were tagged. This procedure took advantage of two subsequent rounds of FACS efficiently selecting the top fluorescent cells within one month (4.1.1, Fig. 4-1). To ensure that the tagged single cell clones mediated stable transgene expression, only those cells were used as master clones whose fluorescence did not vary over at least three months. Depending on the architecture of the GFP tagging vector, the reporter was either excised or exchanged by another transgene.

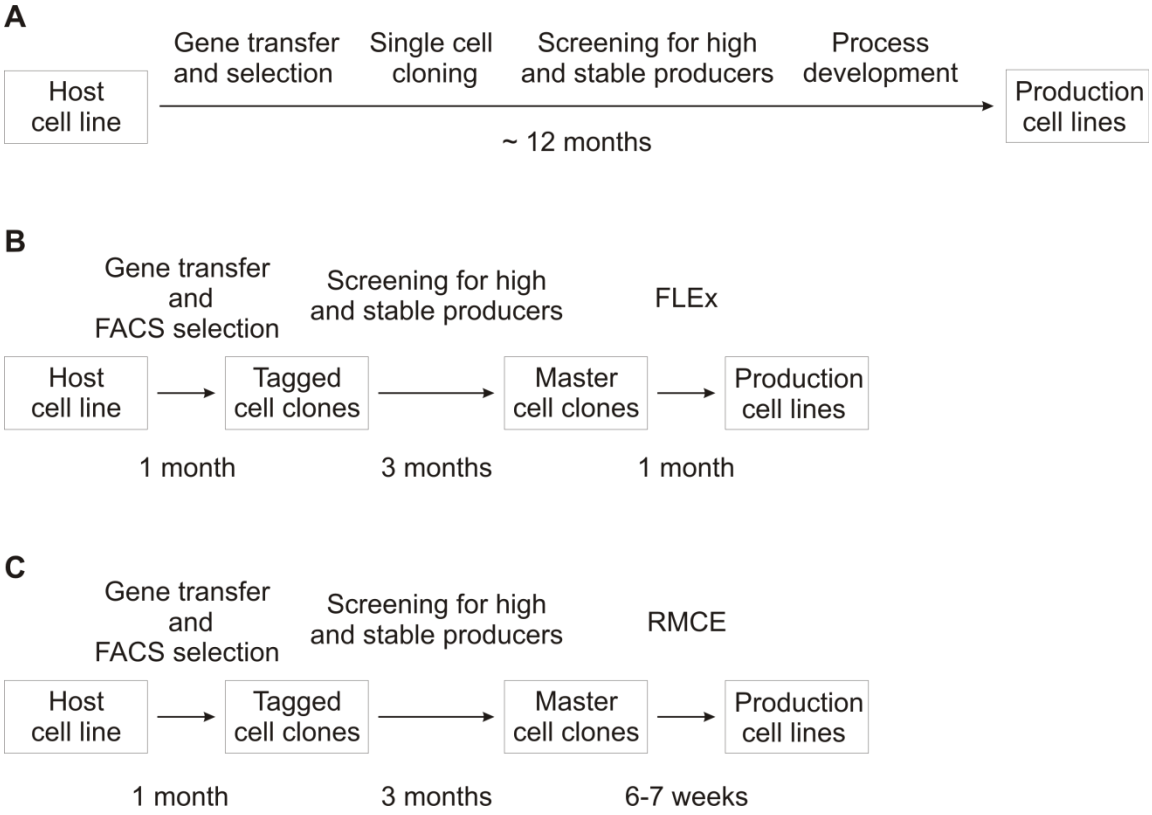


Figure 4-1: Timeline comparison of different cell line development strategies. (A) The conventional strategy needs a minimum of 12 months from gene transfer over screening for high and stable producer cell clones to develop large-scale production processes (Wurm, 2004). (B) The FLEx strategy relies on GFP expressing master cell clones that were selected by two rounds of FACS (1 month). The selected high producers were screened for their stability (3 months) and the best master cell clones were subjected to Flp-mediated recombination resulting in the excision of GFP and activation of the desired transgene. First cryo-preserved production cell clones were available after another subcell cloning step (1 month). Altogether 5 months were needed for this strategy. This could be shortened to 4 months, when the FLEx reaction is performed in parallel to the last weeks of screening. (C) The generation of the RMCE master cells is analogous to the FLEx system and took 4 months. The RMCE reaction exchanging the GFP gene against another transgene resulted in the production cell lines. This step took 6 weeks for scHGF subclones and 7 weeks for RFP subclones depending on the number of subcloning rounds. Altogether the whole process took about 5.5 months for the first production cell line and 6 to 7 weeks for the following ones, as the master cell clones could be reused.

While FLEx is a monomolecular reaction and therefore thermodynamically favoured, RMCE is a rare event. Recombination frequencies were up to 36% for FLEx and only up to 0.003% for RMCE. Thus, RMCE needed a stringent selection by the Δneo^R selection trap to isolate the recombined cell clones (4.1.2). The additional antibiotic selection explains why RMCE took six to seven weeks, while the FLEx step took only one month and was easier to establish. Moreover, the latter strategy completely circumvented the use of antibiotic selection. According to published data, the prokaryotic neomycin resistance gene acts as a transcriptional silencer in eukaryotic cells (Artelt *et al.*, 1991). However, this was not observed in this work. To minimize the risk of instabilities, a

combined use of RMCE and FLEx would be conceivable that would eliminate the selection trap after successful RMCE.

An important factor of cell line development by FACS and recombination strategies is the short and predictable timeline. Dependent on the strategy the first cryo-preserved cells were available after 4 months and 5.5 months for the FLEx and the RMCE strategy, respectively (Fig. 4-1). This is a big advantage compared to conventional cell line development requiring about 12 months (Wurm, 2004). The RMCE system is even more convenient, as the master cells are universal and only have to be established once. They can be converted to express any desired transgene, which shortened the timeline further. The expansion of first subclones started three weeks post transfection, followed by cryo-preservation after six to seven weeks depending on the number of subcloning cycles. Altogether, 10 different production cell lines were rapidly established from selected master cells. The recombinant products include glycoproteins, an intracellular protein and a GPI-anchored protein.

The RMCE cell lines for scHGF and hLAMP3-prox were cultivated in parallel to their FLEx counterparts in 2.5 L bench top bioreactors to compare their productivities. It is worth noting that the scale up from 6 well plates to 2.5 L was a straightforward process for all stable cell lines using a standard protocol for the bioreactor process. In contrast, the scale up for transient transfection approaches is currently still connected to laborious optimization and the preparation of large DNA amounts. Thus, stable cell lines are the favoured expression system especially for low yield glycoproteins that have to be purified from larger culture volumes.

The purified hLAMP3-prox yields derived from the RMCE and the FLEx system were 1 mg and 0.4 mg per litre cell culture, respectively. A direct comparison of scHGF titres in the culture supernatants also resulted in a higher specific productivity for the RMCE cells (1.4 pcd) than for the FLEx cells (1 pcd). The measured productivities are in accordance with the relative GFP fluorescence levels of the respective master cell clones.

In summary, RMCE with a suitable, well characterized master cell line was found to be faster and more predictable than FLEx. Using RMCE, fewer clones had to be analysed

and preparative cell sorting equipment was not required after the master cell clones had been established. Due to these facts, RMCE was an ideal system to produce mammalian glycoproteins.

4.2 Production of glycoproteins for X-ray crystallography

4.2.1 The glycosylation dilemma

Structural information of mammalian glycoproteins gained an increased attention in recent years, as they often have key functions in the human body. However, their production is challenging and requires eukaryotic expression systems though being cost and time intensive. Moreover, crystallography projects often necessitate the production of a variety of different constructs per protein to find well crystallisable protein. As discussed above, the RMCE system in CHO cells is an excellent choice to enhance mammalian cell line development for glycoprotein production.

Another challenge for X-ray structure analysis is the heterogeneous and flexible nature of mammalian protein glycosylation which often hampers the formation of well diffracting crystals (Butters *et al.*, 1999). In contrast, the production of completely deglycosylated glycoprotein may lead to folding problems and instable products. Glycosylation mutant cell lines like CHO Lec3.2.8.1 are a good alternative producing glycoproteins with truncated carbohydrate chains of high-mannose type (Stanley, 1989). Moreover, they have the advantage of the mammalian machinery for protein folding and processing. Yeasts also synthesise glycans of the high-mannose type, but fail to efficiently secrete certain classes of mammalian proteins (Chang *et al.*, 2007). If glycoproteins with high-mannose type side chains remain not crystallisable, a further deglycosylation often results in improvement (Chang *et al.*, 2007).

4.2.2 CHO Lec3.2.8.1 cells are beneficial for X-ray crystallography

In this work, two glycoprotein constructs, scHGF and hLAMP3-prox, were analysed regarding their glycosylation homogeneity. They were produced in glycosylation mutant CHO Lec3.2.8.1 cells upon RMCE. The production process could be kept quite simple, as no additional inhibitors of the glycosylation processing pathway, like NB-DNJ, swainosonine or kifunensine, had to be added to gain shorter and more homogenous glycosylation. The high-mannose type glycans added by CHO Lec3.2.8.1 cells

(Fig. 1-10) can be efficiently digested by endoglycosidase H (endo H). Endo H cleaves between GlcNAc residues in the di-N-acetyl-chitobiose core of high mannose and hybrid-type N-glycans, leaving single GlcNAc units at each glycosylation site (Chang *et al.*, 2007). The single GlcNAc is known to shield hydrophobic regions of the protein surface (Petrescu *et al.*, 2004). Proteins bearing these GlcNAc residues were found to be more stable than those completely deglycosylated by PNGase F (Davis *et al.*, 1995). Deglycosylation of purified scHGF with 30 U/ µg endo H at pH 5.2 followed by mass spectrometry indicated that three of the four predicted glycosylation sites were occupied. The positions were identified as N294, N566 and N653, all bearing $GlcNAc_2Man_5$ chains. N402 was found non-glycosylated. That proved the high homogeneity of glycoproteins concerning glycan composition and occupied sites when using CHO Lec3.2.8.1 cells.

In case of pure hLAMP3-prox, however, non-uniformly occupied glycosylation sites were identified. One (predominantly N291), two (predominantly N266 and N291) or all three sites (N232, N266 and N291) of the protein bore $GlcNAc_2Man_{4-5}$ chains. Total hLAMP3-prox recovery following endo H digestion was higher at pH 7.4 than pH 5.2, signalling pH sensitivity of the deglycosylated protein. Nevertheless, all one to three occupied glycosylation sites were efficiently shortened to one GlcNAc residue conferring a suitable homogeneity for crystallisation (4.3.2). Before crystallisation, endo H was separated from hLAMP3-prox by GPC. The deglycosylated protein was soluble for at least two weeks. As the stability of glycosylated hLAMP3-prox was even higher, deglycosylation was applied in several batches depending on the need for crystallisation. Digestion by endo H can be easily adapted to different scales.

Taken together, these observations confirm that glycoproteins produced by CHO Lec3.2.8.1 cells are well susceptible for endo H digestion resulting in soluble homogenous protein with a minimal amount of flexible sugars. Thus, these hosts are an excellent expression system for X-ray crystallography.

4.3 Structural analysis of hLAMP3-prox

LAMP-3 belongs to the family of lysosome accociated glycoproteins. It was suggested that LAMP-3 is involved in processing of exogenous antigens restricted to MHC class II presentation, as it is associated with the intracellular MHC class II compartment in late

mature dendritic cells (de Saint-Vis *et al.*, 1998). LAMP-3 may facilitate the translocation of MHC class II molecules to the cell surface. However, direct information about the function of LAMP-3 is still awaited, for example by generation of knock out animals.
The sections below discuss the progress towards a crystal structure of the membrane-proximal domain of human LAMP-3 in lysosomes. This structure is the first among all LAMP family members. Production, purification and crystallisation conditions were established. The initial model from the crystal structure was finally determined using multi-wavelength anomalous dispersion. It revealed a β-sandwich fold of two β-sheets.

4.3.1 Production and purification of hLAMP3-prox

Since hLAMP3-prox is a mammalian glycoprotein with three predicted N-glycosylation sites, it was produced in CHO Lec3.2.8.1 cell lines secreting the protein with a C-terminal hexa-histidine (His_6) tag for purification. The production cell lines were established by FLEx and RMCE as described before in this work. Perfusion processes routinely resulted in 22.5 L conditioned medium which was concentrated to 1 L followed by a buffer change and IMAC purification. The affinity chromatography should specifically purify the tagged protein and should concentrate it further upon elution. However, this step was not as specific as expected. Many proteins secreted by CHO cells were unspecifically bound to the Ni-NTA column and eluted together with hLAMP3-prox. This phenomenon was also observed when purifying other LAMP constructs by the same protocol (Dr. Konrad Büssow, personal communication). In case of hLAMP3-prox the contaminants were removed by GPC due to sufficient size differences. This two-step procedure yielded pure monodisperse protein which could be concentrated to at least 10 mg/ mL. As unspecific bound proteins in the culture supernatant of CHO Lec3.2.8.1 cells seem to be a general problem for purifying His_6 tagged proteins, alternative strategies are discussed in the following. One possibility would be the increase in length of the polyhistidine tags from His_6 to His_{10} which usually coincides with increased recovery and purity, but may in some cases reduce the expression (Grisshammer & Tucker, 1997; Mohanty & Wiener, 2004). Alternatively, other tags can be added at either terminus for an initial affinity chromatography step. They include for example Strep II tag, OneSTrEP, Flag-tag, GST-tag, MBP-tag, Fc-tag, and myc-tag. These tags differ substantially in purity, yield and cost (Lichty *et al.*, 2005). The choice of the tag depends on the requirement of an experiment. A comparative study using a tagged dihydrofolate reductase (DHFR) construct produced in HeLa cells revealed that

GST tags as well as His tags are favoured for low-cost purification of large quantities of partial purified protein (Lichty *et al.*, 2005). The highest purity was obtained by epitope tags such as the FLAG-tag, though being a cost intensive method. The study further found the Strep II tag to be an excellent candidate for affinity purification of DHFR, since it is a short tag resulting in pure protein in good yields and at a moderate cost.

For structural studies it may be important to remove large affinity tags like GST or MBP after purification to obtain natural protein without unstructured, flexible residues. This can be realized by integrating a protease recognition sequence. It has been shown that TEV protease cleavage was efficient to remove N-terminal MBP and C-terminal thioredoxin-His_6 tags for example from the neurotoxin receptor (White *et al.*, 2004). It should be noted that polyhistidine tags are often fused to the C-terminus of a recombinant protein to maintain expression (Massotte, 2003) and that protease recognition sites generally leave between four and six additional residues behind upon cleavage when placed at the C-terminus. In case of hLAMP3-prox, the domain is naturally connected to a trans-membrane domain at its C-terminus, and thus the C-terminal His_6-tag was not expected to alter the fold of the construct.

Crystallisation approaches with glycosylated hLAMP3-prox failed (data not shown). Therefore, the protein was deglycosylated by endo H. This procedure included another GPC purification, followed by concentrating up to 22 mg/ mL without any noticeable precipitation at room temperature. DLS measurements proved the monodispersity before crystallisation.

4.3.2 First insights into X-ray structure of hLAMP3-prox

After purification and crystallisation of hLAMP3-prox, its initial structure was obtained to a resolution of 2.8 Å by MAD. To this end, a screening for heavy atom derivatization was required which was performed by native SDS-PAGE. This technique successfully identified a compound for phasing and thus represents a practical alternative to selenomethionine labelling, which was impaired in this study by the toxicity of selenomethionine to CHO Lec3.2.8.1 cells (data not shown).

The model of the membrane-proximal luminal domain of LAMP-3 provided first insights into its previously unknown three-dimensional structure. Only seven amino acid

residues near the N-terminus could not be located, possibly due to their high flexibility (Fig. 3-32). Further attempts to improve the electron density map and thus the initial model might include screening for crystals with higher resolution.

In accordance with secondary structure prediction, the initial model reveals a high β-fold content with two anti-parallel β-sheets forming a sandwich. Comparative structural analysis using the DALI-Server did not find any convincing homologue. This indicates that the hLAMP3-prox domain contains new structural features. Interestingly, the C- and N-terminus of the crystallised domain are located in close proximity (Fig. 3-32 B). In the full length protein, the C-terminus of the membrane-proximal domain is connected to a transmembrane helix followed by a short cytoplasmic domain. The N-terminus is connected to a serine/proline-rich region followed by a highly glycosylated mucin-like domain. Hence, the mucin-like domain seems to be located near the lysosomal membrane and may protect the membrane and other membrane proteins from the acidic conditions of the lumen.

4.3.3 GlcNAc-stabilized crystal packing

Two of three expected GlcNAc residues (N266 and N291) were visible in the electron density map of hLAM3-prox. The third glycosylation site at position N232 was located in the flexible loop region and thus could not be built into the model. The GlcNAc residue connected to N291 stabilized the crystal packing by contacting the crystal neighbour's backbone at position G287. Crystal contacts by single GlcNAc residues have been reported before (Petrescu *et al.*, 2004). This feature might explain why hLAMP3-prox bearing high-mannose glycans did not crystallize in this condition.

4.3.4 Structure of hLAMP3-prox as phasing model

The LAMP membrane-proximal domain is the most conserved region among all LAMP family members. Separate amino acid alignments of this domain showed that human LAMP-3 shares 31% sequence identity with CD68, 27% identity with LAMP-1 and 25% identity with LAMP-2 (de Saint-Vis *et al.*, 1998). It would be worth trying to use hLAMP3-prox as a phasing model to determine the domain structures of other LAMP family members, although the success rate of molecular replacement decreases sharply when the sequence similarity between template and target proteins drops below 40% identical

residues. However, there are examples of structures solved using molecular replacement with phasing models of less than 20% sequence identity (Jones, 2001).

4.4 Outlook

The combination of FACS and Flp-recombination for cell line development in CHO Lec3.2.8.1 cells has improved the production of glycoproteins for X-ray crystallography. RMCE paved the way for structure determination of hLAMP3-prox in a minimal amount of time and was used to establish a number of other cell lines stably expressing protein in mg-per-litre scale with less effort. The system is currently being used to facilitate the production of further recombinant glycoproteins for their structural analysis. Thus, it has the potential to standardize the glycoprotein production in mammalian cells. Moreover, the described strategy is not limited to the glycosylation mutant cell line, but can also be applied on other cell lines like CHO K1, BHK, HEK293 or PER.C6 commonly used in pharmaindustry to produce therapeutic proteins.

A possible advancement would be the parallel expression of different recombinant proteins in the same cell for assembling of multi-protein complexes, as already discussed in 4.1.5. Besides the establishment of a multiplex system (Turan *et al.*, 2010), another strategy would be a simultaneous cultivation of a subclone mixture producing different glycoproteins. The recombinant protein complexes could assemble in the culture supernatant and thus could be co-purified. If the cultivated subclones derived from the same master cell clone by RMCE, they should be isogenic except for the transgenes. Hence, the result would be a heterogeneous production culture with homogenous cultivation abilities.

To allow for even stronger expression of challenging proteins, the presented RMCE system has potential for further enhancements. One approach might aim at transcriptional augmentation of hard to express proteins by RMCE. This might be achieved by integrating cis-acting DNA sequences. The addition of S/MARs at the borders of a target cassette resulted in a larger proportion of master cells with high reporter gene expression (Qiao *et al.*, 2009). This indicated that S/MARs alleviated silencing and increased transfection-mediated integration events. Due to the sorting strategy for master cell selection, a larger proportion of high producers is not necessarily needed. However, based on their DNA strand separation potential, it is

further argued that S/MARs support recombination events like RMCE (Qiao *et al.*, 2009). Another study suggested, however, that S/MARs integrated upon RMCE at a weak expression locus was not able to significantly improve the expression level (Nehlsen *et al.*, 2009). Thus, the value of S/MAR elements might be more pronounced in increasing low recombination frequencies.

Another strategy to increase the expression level of hard-to-express proteins is metabolic engineering by using multiplex RMCE. Thereby, positive effects of the co-expression of transcriptional activators, anti-apoptotic proteins or secretion supporting proteins can be analysed individually and more easily. Former studies revealed increased expression levels by co-expression of the transcription activator XBP-1 (Becker *et al.*, 2008), the caspase inhibitor XIAP (Sauerwald *et al.*, 2002) or the modulator of protein secretion CERT (Florin *et al.*, 2009).

Unfortunately, some structurally interesting proteins are toxic for the host cells in high concentrations. Thus, a regulated recombinant expression is desired and can be achieved by using inducible promoter systems. Commonly used is a tetracycline-regulated gene expression system (tet system) (Corbel & Rossi, 2002; Gossen & Bujard, 1992). Basically, a repressor binds to a tet operator within a modified promoter and prevents the assembling of the transcription initiation complex. In the presence of the antibiotic tetracycline, the repressor dissociates and the transcription is induced. Depending on the tetracycline concentration in the culture medium, even the expression strength can be regulated to minimize deleterious effects of cytotoxic proteins.

Insect cells are another excellent expression system to produce proteins for structural studies (Nettleship *et al.*, 2010). Establishing the RMCE system in these cells would be beneficial. By designing exchange vectors that can be used in both hosts, selection of the best suited expression system would be facilitated. Generally, cytosolic proteins are favourably produced in insect cells, while for secreted proteins mammalian cells are preferred (Aricescu *et al.*, 2006). Hence, a RMCE system for both expression hosts would be complementary for efficient protein production.

In conclusion, RMCE in mammalian cells can overcome the production bottleneck for glycoproteins. The established system can be routinely used by structural biologists to

enhance determination of three-dimensional glycoprotein structures. Time that is saved for the production step can accelerate the functional analysis of proteins involved in pathogenic pathways and thus drug discovery.

References

Abremski, K., Hoess, R. & Sternberg, N. (1983). Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. *Cell* **32**, 1301-11.

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21.

Akasaki, K., Nakamura, N., Tsukui, N., Yokota, S., Murata, S., Katoh, R., Michihara, A., Tsuji, H., Marques, E. T., Jr. & August, J. T. (2004). Human dendritic cell lysosome-associated membrane protein expressed in lung type II pneumocytes. *Arch Biochem Biophys* **425**, 147-57.

Andersen, D. C. & Krummen, L. (2002). Recombinant protein expression for therapeutic applications. *Curr Opin Biotechnol* **13**, 117-23.

Antoniou, M., Harland, L., Mustoe, T., Williams, S., Holdstock, J., Yague, E., Mulcahy, T., Griffiths, M., Edwards, S., Ioannou, P. A., Mountain, A. & Crombie, R. (2003). Transgenes encompassing dual-promoter CpG islands from the human TBP and HNRPA2B1 loci are resistant to heterochromatin-mediated silencing. *Genomics* **82**, 269-79.

Apweiler, R., Hermjakob, H. & Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**, 4-8.

Aricescu, A. R., Assenberg, R., Bill, R. M., Busso, D., Chang, V. T., Davis, S. J., Dubrovsky, A., Gustafsson, L., Hedfalk, K., Heinemann, U., Jones, I. M., Ksiazek, D., Lang, C., Maskos, K., Messerschmidt, A., Macieira, S., Peleg, Y., Perrakis, A., Poterszman, A., Schneider, G., Sixma, T. K., Sussman, J. L., Sutton, G., Tarboureich, N., Zeev-Ben-Mordehai, T. & Jones, E. Y. (2006). Eukaryotic expression: developments for structural proteomics. *Acta Crystallogr D Biol Crystallogr* **62**, 1114-24.

Artelt, P., Grannemann, R., Stocking, C., Friel, J., Bartsch, J. & Hauser, H. (1991). The prokaryotic neomycin-resistance-encoding gene acts as a transcriptional silencer in eukaryotic cells. *Gene* **99**, 249-54.

Ausubel, F. M., Brent, R., Kingston, R. E., Moor, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (2007). *Current protocols in molecular biology.*, John Wiley and Sons Inc., New York.

Baer, A. & Bode, J. (2001). Coping with kinetic and thermodynamic barriers: RMCE, an efficient strategy for the targeted integration of transgenes. *Curr Opin Biotechnol* **12**, 473-80.

Baldi, L., Hacker, D. L., Adam, M. & Wurm, F. M. (2007). Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. *Biotechnol Lett* **29**, 677-84.

Barnes, L. M., Bentley, C. M. & Dickson, A. J. (2003). Stability of protein production from recombinant mammalian cells. *Biotechnol Bioeng* **81**, 631-9.

Becker, E., Florin, L., Pfizenmaier, K. & Kaufmann, H. (2008). An XBP-1 dependent bottle-neck in production of IgG subtype antibodies in chemically defined serum-free Chinese hamster ovary (CHO) fed-batch processes. *J Biotechnol* **135**, 217-23.

Becker, E., Florin, L., Pfizenmaier, K. & Kaufmann, H. (2010). Evaluation of a combinatorial cell engineering approach to overcome apoptotic effects in XBP-1(s) expressing cells. *J Biotechnol* **146**, 198-206.

Benton, T., Chen, T., McEntee, M., Fox, B., King, D., Crombie, R., Thomas, T. C. & Bebbington, C. (2002). The use of UCOE vectors in combination with a preadapted serum free, suspension cell line allows for rapid production of large quantities of protein. *Cytotechnology* **38**, 43-6.

Bestor, T. H. (2000). Gene silencing as a threat to the success of gene therapy. *J Clin Invest* **105**, 409-11.

Bijvoet, J. M. (1954). Structure of Optically Active Compounds in the Solid State. *Nature* **173**, 888-891.

Birchmeier, C., Birchmeier, W., Gherardi, E. & Vande Woude, G. F. (2003). Met, metastasis, motility and more. *Nat Rev Mol Cell Biol* **4**, 915-25.

Birnboim, H. C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**, 1513-23.

Bishop, B., Aricescu, A. R., Harlos, K., O'Callaghan, C. A., Jones, E. Y. & Siebold, C. (2009). Structural insights into hedgehog ligand sequestration by the human hedgehog-interacting protein HHIP. *Nat Struct Mol Biol* **16**, 698-703.

Bode, J., Schlake, T., Iber, M., Schubeler, D., Seibler, J., Snezhkov, E. & Nikolaev, L. (2000). The transgeneticist's toolbox: novel methods for the targeted modification of eukaryotic genomes. *Biol Chem* **381**, 801-13.

Boggon, T. J. & Shapiro, L. (2000). Screening for phasing atoms in protein crystallography. *Structure* **8**, R143-9.

Bollati-Fogolin, M., Forno, G., Nimtz, M., Conradt, H. S., Etcheverrigaray, M. & Kratje, R. (2005). Temperature reduction in cultures of hGM-CSF-expressing CHO cells: effect on productivity and product quality. *Biotechnol Prog* **21**, 17-21.

Borth, N., Mattanovich, D., Kunert, R. & Katinger, H. (2005). Effect of increased expression of protein disulfide isomerase and heavy chain binding protein on antibody secretion in a recombinant CHO cell line. *Biotechnol Prog* **21**, 106-11.

Bowden, T. A., Aricescu, A. R., Gilbert, R. J., Grimes, J. M., Jones, E. Y. & Stuart, D. I. (2008a). Structural basis of Nipah and Hendra virus attachment to their cell-surface receptor ephrin-B2. *Nat Struct Mol Biol* **15**, 567-72.

Bowden, T. A., Crispin, M., Harvey, D. J., Aricescu, A. R., Grimes, J. M., Jones, E. Y. & Stuart, D. I. (2008b). Crystal structure and carbohydrate analysis of Nipah virus attachment glycoprotein: a template for antiviral and vaccine design. *J Virol* **82**, 11628-36.

Buchanan, S. G. (2002). Structural genomics: bridging functional genomics and structure-based drug design. *Curr Opin Drug Discov Devel* **5**, 367-81.

Buchholz, F., Angrand, P. O. & Stewart, A. F. (1996a). A simple assay to determine the functionality of Cre or FLP recombination targets in genomic manipulation constructs. *Nucleic Acids Res* **24**, 3118-9.

Buchholz, F., Angrand, P. O. & Stewart, A. F. (1998). Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat Biotechnol* **16**, 657-62.

Buchholz, F., Ringrose, L., Angrand, P. O., Rossi, F. & Stewart, A. F. (1996b). Different thermostabilities of FLP and Cre recombinases: implications for applied site-specific recombination. *Nucleic Acids Res* **24**, 4256-62.

Butters, T. D., Sparks, L. M., Harlos, K., Ikemizu, S., Stuart, D. I., Jones, E. Y. & Davis, S. J. (1999). Effects of N-butyldeoxynojirimycin and the Lec3.2.8.1 mutant phenotype on N-glycan processing in Chinese hamster ovary cells: application to glycoprotein crystallization. *Protein Sci* **8**, 1696-701.

Cachianes, G., Ho, C., Weber, R. F., Williams, S. R., Goeddel, D. V. & Leung, D. W. (1993). Epstein-Barr virus-derived vectors for transient and stable expression of recombinant proteins. *Biotechniques* **15**, 255-9.

Carafoli, F., Clout, N. J. & Hohenester, E. (2009). Crystal structure of the LG1-3 region of the laminin alpha2 chain. *J Biol Chem* **284**, 22786-92.

Carafoli, F., Saffell, J. L. & Hohenester, E. (2008). Structure of the tandem fibronectin type 3 domains of neural cell adhesion molecule. *J Mol Biol* **377**, 524-34.

Caron, A. W., Nicolas, C., Gaillet, B., Ba, I., Pinard, M., Garnier, A., Massie, B. & Gilbert, R. (2009). Fluorescent labeling in semi-solid medium for selection of mammalian cells secreting high-levels of recombinant proteins. *BMC Biotechnol* **9**, 42.

CCPN4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50 (Pt 5)**, 760-763.

Chang, V. T., Crispin, M., Aricescu, A. R., Harvey, D. J., Nettleship, J. E., Fennelly, J. A., Yu, C., Boles, K. S., Evans, E. J., Stuart, D. I., Dwek, R. A., Jones, E. Y., Owens, R. J. & Davis, S. J. (2007). Glycoprotein Structural Genomics: Solving the Glycosylation Problem. *Structure* **15**, 267-273.

Chen, Y. & Rice, P. A. (2003). New insight into site-specific recombination from Flp recombinase-DNA structures. *Annu Rev Biophys Biomol Struct* **32**, 135-59.

Chmielowiec, J., Borowiak, M., Morkel, M., Stradal, T., Munz, B., Werner, S., Wehland, J., Birchmeier, C. & Birchmeier, W. (2007). c-Met is essential for wound healing in the skin. *J Cell Biol* **177**, 151-62.

Cobellis, G., Nicolaus, G., Iovino, M., Romito, A., Marra, E., Barbarisi, M., Sardiello, M., Di Giorgio, F. P., Iovino, N., Zollo, M., Ballabio, A. & Cortese, R. (2005). Tagging genes with cassette-exchange sites. *Nucleic Acids Res* **33**, e44.

Cockett, M. I., Bebbington, C. R. & Yarranton, G. T. (1990). High level expression of tissue inhibitor of metalloproteinases in Chinese hamster ovary cells using glutamine synthetase gene amplification. *Biotechnology (N Y)* **8**, 662-7.

Cole, C., Barber, J. D. & Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* **36**, W197-201.

Coligan, J. E., Dunn, B. M., Ploegh, H. L., Speicher, D. W. & Wingfield, P. T. (2002). *Current protocols in protein science.*, John Wiley and Sons Inc., New York.

Corbel, S. Y. & Rossi, F. M. (2002). Latest developments and in vivo use of the Tet system: ex vivo and in vivo delivery of tetracycline-regulated genes. *Curr Opin Biotechnol* **13**, 448-52.

Crispin, M., Bowden, T. A., Coles, C. H., Harlos, K., Aricescu, A. R., Harvey, D. J., Stuart, D. I. & Jones, E. Y. (2009). Carbohydrate and domain architecture of an immature antibody glycoform exhibiting enhanced effector functions. *J Mol Biol* **387**, 1061-6.

Dauter, Z., Dauter, M. & Dodson, E. (2002). Jolly SAD. *Acta Crystallogr D Biol Crystallogr* **58**, 494-506.

Davie, J. R. (2003). Inhibition of histone deacetylase activity by butyrate. *J Nutr* **133**, 2485S-2493S.

Davis, R., Schooley, K., Rasmussen, B., Thomas, J. & Reddy, P. (2000). Effect of PDI overexpression on recombinant protein secretion in CHO cells. *Biotechnol Prog* **16**, 736-43.

Davis, S. J., Davies, E. A., Barclay, A. N., Daenke, S., Bodian, D. L., Jones, E. Y., Stuart, D. I., Butters, T. D., Dwek, R. A. & van der Merwe, P. A. (1995). Ligand binding by the immunoglobulin superfamily recognition molecule CD2 is glycosylation-independent. *J Biol Chem* **270**, 369-75.

Davis, S. J., Puklavec, M. J., Ashford, D. A., Harlos, K., Jones, E. Y., Stuart, D. I. & Williams, A. F. (1993). Expression of soluble recombinant glycoproteins with predefined glycosylation: application to the crystallization of the T-cell glycoprotein CD2. *Protein Eng* **6**, 229-32.

de Saint-Vis, B., Vincent, J., Vandenabeele, S., Vanbervliet, B., Pin, J. J., Ait-Yahia, S., Patel, S., Mattei, M. G., Banchereau, J., Zurawski, S., Davoust, J., Caux, C. & Lebecque, S. (1998). A novel lysosome-associated membrane glycoprotein, DC-LAMP, induced upon DC maturation, is transiently expressed in MHC class II compartment. *Immunity* **9**, 325-36.

DeKelver, R. C., Choi, V. M., Moehle, E. A., Paschon, D. E., Hockemeyer, D., Meijsing, S. H., Sancak, Y., Cui, X., Steine, E. J., Miller, J. C., Tam, P., Bartsevich, V. V., Meng, X., Rupniewski, I., Gopalan, S. M., Sun, H. C., Pitz, K. J., Rock, J. M., Zhang, L., Davis, G. D., Rebar, E. J., Cheeseman, I. M., Yamamoto, K. R., Sabatini, D. M., Jaenisch, R., Gregory, P. D. & Urnov, F. D. (2010). Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res* **20**, 1133-42.

Derouazi, M., Girard, P., Van Tilborgh, F., Iglesias, K., Muller, N., Bertschinger, M. & Wurm, F. M. (2004). Serum-free large-scale transient transfection of CHO cells. *Biotechnol Bioeng* **87**, 537-45.

Derouazi, M., Martinet, D., Besuchet Schmutz, N., Flaction, R., Wicht, M., Bertschinger, M., Hacker, D. L., Beckmann, J. S. & Wurm, F. M. (2006). Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. *Biochem Biophys Res Commun* **340**, 1069-77.

Duda, A. (2009). Charakterisierung von Flp Rekombinations-Systemen in Säugerzelllinien. Diploma thesis.

Durocher, Y., Perret, S. & Kamen, A. (2002). High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res* **30**, E9.

Eder, J. P., Vande Woude, G. F., Boerner, S. A. & LoRusso, P. M. (2009). Novel therapeutic inhibitors of the c-Met signaling pathway in cancer. *Clin Cancer Res* **15**, 2207-14.

Edman, P. & Begg, G. (1967). A protein sequenator. *Eur J Biochem* **1**, 80-91.

Emerman, M. & Temin, H. M. (1984). Genes with promoters in retrovirus vectors can be independently suppressed by an epigenetic mechanism. *Cell* **39**, 449-67.

Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32.

Eskelinen, E. L. & Saftig, P. (2009). Autophagy: a lysosomal degradation pathway with a central role in health and disease. *Biochim Biophys Acta* **1793**, 664-73.

Evans, G. & Pettifer, R. F. (2001). CHOOHC: a program for deriving anomalous scattering factors from x-ray fluorescence spectra. *J Appl Chrystallogr* **34**, 82-86.

Florin, L., Pegel, A., Becker, E., Hausser, A., Olayioye, M. A. & Kaufmann, H. (2009). Heterologous expression of the lipid transfer protein CERT increases therapeutic protein productivity of mammalian cells. *J Biotechnol* **141**, 84-90.

Francastel, C., Walters, M. C., Groudine, M. & Martin, D. I. (1999). A functional enhancer suppresses silencing of a transgene and prevents its localization close to centrometric heterochromatin. *Cell* **99**, 259-69.

Friedel, G. (1913). Sur les symétries cristallines que peut révéler la diffraction des rayons X. *C.R. Acad. Sci. Paris* **157**, 1533-1536.

Funakoshi, H. & Nakamura, T. (2003). Hepatocyte growth factor: from diagnosis to clinical applications. *Clin Chim Acta* **327**, 1-23.

Gaines, P. & Wojchowski, D. M. (1999). pIRES-CD4t, a dicistronic expression vector for MACS- or FACS-based selection of transfected cells. *Biotechniques* **26**, 683-8.

Gellissen, G. (2005). Production of recombinant proteins - Novel microbial and eukaryotic expression systems. (Gellissen, G., ed.). Wiley-VCH, Weinheim.

Gherardi, E., Sandin, S., Petoukhov, M. V., Finch, J., Youles, M. E., Ofverstedt, L. G., Miguel, R. N., Blundell, T. L., Vande Woude, G. F., Skoglund, U. & Svergun, D. I. (2006). Structural basis of hepatocyte growth factor/scatter factor and MET signalling. *Proc Natl Acad Sci U S A* **103**, 4046-51.

Girard, P., Derouazi, M., Baumgartner, G., Bourgeois, M., Jordan, M., Jacko, B. & Wurm, F. M. (2002). 100-liter transient transfection. *Cytotechnology* **38**, 15-21.

Girod, P. A., Zahn-Zabal, M. & Mermod, N. (2005). Use of the chicken lysozyme 5' matrix attachment region to generate high producer CHO cell lines. *Biotechnol Bioeng* **91**, 1-11.

Glaser, S., Anastassiadis, K. & Stewart, A. F. (2005). Current issues in mouse genome engineering. *Nat Genet* **37**, 1187-93.

Gossen, M. & Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* **89**, 5547-51.

Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol* **36**, 59-74.

Grisshammer, R. & Tucker, J. (1997). Quantitative evaluation of neurotensin receptor purification by immobilized metal affinity chromatography. *Protein Expr Purif* **11**, 53-60.

Gurtu, V., Yan, G. & Zhang, G. (1996). IRES bicistronic expression vectors for efficient creation of stable mammalian cell lines. *Biochem Biophys Res Commun* **229**, 295-8.

Hahn, T. (1993). *International Tables for Crystallography* (Hahn, T., Ed.), A, A.D. Reidel.

Harker, D. (1936). The application of the three-dimensional patterson method and the crystal structures of proustite, Ag_3AsS_3, and pyrargyrite, Ag_3SbS_3. *J. Chem. Phys.* **4**, 381-390.

Hoeijmakers, J. H., Odijk, H. & Westerveld, A. (1987). Differences between rodent and human cell lines in the amount of integrated DNA after transfection. *Exp Cell Res* **169**, 111-9.

Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. (2008). Searching protein structure databases with DaliLite v.3. *Bioinformatics* **24**, 2780-1.

Holm, L. & Sander, C. (1999). Protein folds and families: sequence and structure alignments. *Nucleic Acids Res* **27**, 244-7.

Holness, C. L. & Simmons, D. L. (1993). Molecular cloning of CD68, a human macrophage marker related to lysosomal glycoproteins. *Blood* **81**, 1607-13.

Huang, Y., Li, Y., Wang, Y. G., Gu, X., Wang, Y. & Shen, B. F. (2007). An efficient and targeted gene integration system for high-level antibody expression. *J Immunol Methods* **322**, 28-39.

Huh, C. G., Factor, V. M., Sanchez, A., Uchida, K., Conner, E. A. & Thorgeirsson, S. S. (2004). Hepatocyte growth factor/c-met signaling pathway is required for efficient liver regeneration and repair. *Proc Natl Acad Sci U S A* **101**, 4477-82.

Huynh, K. K., Eskelinen, E. L., Scott, C. C., Malevanets, A., Saftig, P. & Grinstein, S. (2007). LAMP proteins are required for fusion of lysosomes with phagosomes. *Embo J* **26**, 313-24.

Jones, D. T. (2001). Evaluating the potential of using fold recognition models for molecular replacement. *Acta Crystallogr D* **57**, 1428-1434.

Kabsch, W. (2010a). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66**, 133-44.

Kabsch, W. (2010b). Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32.

Kalwy, S., Rance, J. & Young, R. (2006). Toward more efficient protein expression: keep the message simple. *Mol Biotechnol* **34**, 151-6.

Kanao, H., Enomoto, T., Kimura, T., Fujita, M., Nakashima, R., Ueda, Y., Ueno, Y., Miyatake, T., Yoshizaki, T., Buzard, G. S., Tanigami, A., Yoshino, K. & Murata, Y. (2005). Overexpression of LAMP3/TSC403/DC-LAMP promotes metastasis in uterine cervical cancer. *Cancer Res* **65**, 8640-5.

Karreman, S., Hauser, H. & Karreman, C. (1996). On the use of double FLP recognition targets (FRTs) in the LTR of retroviruses for the construction of high producer cell lines. *Nucleic Acids Res* **24**, 1616-24.

Kaufman, W. L., Kocman, I., Agrawal, V., Rahn, H. P., Besser, D. & Gossen, M. (2008). Homogeneity and persistence of transgene expression by omitting antibiotic selection in cell line isolation. *Nucleic Acids Res* **36**, e111.

Kaufmann, H., Mazur, X., Fussenegger, M. & Bailey, J. E. (1999). Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells. *Biotechnol and Bioeng* **63**, 573-582.

Kim, H., Laudemann, J., Stevens, J. & Wu, M. (2009). Expression vector engineering for recombinant protein production. In *Cell line development* (Al-Rubeai, M., ed.), Vol. 6, pp. 97-108. Springer, Dublin.

Kim, J. M., Kim, J. S., Park, D. H., Kang, H. S., Yoon, J., Baek, K. & Yoon, Y. (2004). Improved recombinant gene expression in CHO cells using matrix attachment regions. *J Biotechnol* **107**, 95-105.

Kim, M. S. & Lee, G. M. (2008). Use of Flp-mediated cassette exchange in the development of a CHO cell line stably producing erythropoietin. *J Microbiol Biotechnol* **18**, 1342-51.

Kim, N. S., Byun, T. H. & Lee, G. M. (2001). Key determinants in the occurrence of clonal variation in humanized antibody expression of cho cells during dihydrofolate reductase mediated gene amplification. *Biotechnol Prog* **17**, 69-75.

Kirchhofer, D., Lipari, M. T., Santell, L., Billeci, K. L., Maun, H. R., Sandoval, W. N., Moran, P., Ridgway, J., Eigenbrot, C. & Lazarus, R. A. (2007). Utilizing the activation mechanism of serine proteases to engineer hepatocyte growth factor into a Met antagonist. *Proc Natl Acad Sci U S A* **104**, 5306-11.

Kito, M., Itami, S., Fukano, Y., Yamana, K. & Shibui, T. (2002). Construction of engineered CHO strains for high-level production of recombinant proteins. *Appl Microbiol Biotechnol* **60**, 442-8.

Kunaparaju, R., Liao, M. & Sunstrom, N. A. (2005). Epi-CHO, an episomal expression system for recombinant protein production in CHO cells. *Biotechnol Bioeng* **91**, 670-7.

Lebkowski, J. S., Clancy, S. & Calos, M. P. (1985). Simian virus 40 replication in adenovirus-transformed human cells antagonizes gene expression. *Nature* **317**, 169-71.

Lee, S. K. & Lee, G. M. (2003). Development of apoptosis-resistant dihydrofolate reductase-deficient Chinese hamster ovary cell line. *Biotechnol Bioeng* **82**, 872-6.

Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4+ESF-EAMCB Newsletter on Protein Crystallography.*

Liang, C. P. & Garrard, W. T. (1999). Targeted linearization of DNA in vivo. *Methods* **17**, 95-103.

Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J. & Tan, S. (2005). Comparison of affinity tags for protein purification. *Protein Expr Purif* **41**, 98-105.

Liu, W., Xiong, Y. & Gossen, M. (2006). Stability and homogeneity of transgene expression in isogenic cells. *J Mol Med* **84**, 57-64.

Loonstra, A., Vooijs, M., Beverloo, H. B., Allak, B. A., van Drunen, E., Kanaar, R., Berns, A. & Jonkers, J. (2001). Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc Natl Acad Sci U S A* **98**, 9209-14.

Ma, C. H., Rowley, P. A., Macieszak, A., Guga, P. & Jayaram, M. (2009). Active site electrostatics protect genome integrity by blocking abortive hydrolysis during DNA recombination. *EMBO J* **28**, 1745-56.

Mancia, F., Patel, S. D., Rajala, M. W., Scherer, P. E., Nemes, A., Schieren, I., Hendrickson, W. A. & Shapiro, L. (2004). Optimization of protein production in mammalian cells with a coexpressed fluorescent marker. *Structure* **12**, 1355-60.

Massotte, D. (2003). G protein-coupled receptor overexpression with the baculovirus-insect cell system: a tool for structural and functional studies. *Biochim Biophys Acta* **1610**, 77-89.

Mastrangelo, A. J., Hardwick, J. M., Zou, S. & Betenbaugh, M. J. (2000). Part II. Overexpression of bcl-2 family members enhances survival of mammalian cells in response to various culture insults. *Biotechnol Bioeng* **67**, 555-64.

Mattanovich, D. & Borth, N. (2006). Applications of cell sorting in biotechnology. *Microb Cell Fact* **5**, 12.

Matthews, B. W. (1968). Solvent content of protein crystals. *J Mol Biol* **33**, 491-7.

McBurney, M. W., Mai, T., Yang, X. & Jardine, K. (2002). Evidence for repeat-induced gene silencing in cultured Mammalian cells: inactivation of tandem repeats of transfected genes. *Exp Cell Res* **274**, 1-8.

Meissner, P., Pick, H., Kulangara, A., Chatellard, P., Friedrich, K. & Wurm, F. M. (2001). Transient gene expression: recombinant protein production with suspension-adapted HEK293-EBNA cells. *Biotechnol Bioeng* **75**, 197-203.

Miyazawa, K., Tsubouchi, H., Naka, D., Takahashi, K., Okigaki, M., Arakaki, N., Nakayama, H., Hirono, S., Sakiyama, O. & et al. (1989). Molecular cloning and sequence analysis of cDNA for human hepatocyte growth factor. *Biochem Biophys Res Commun* **163**, 967-73.

Moehle, E. A., Rock, J. M., Lee, Y. L., Jouvenot, Y., DeKelver, R. C., Gregory, P. D., Urnov, F. D. & Holmes, M. C. (2007). Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci U S A* **104**, 3055-60.

Mohanty, A. K. & Wiener, M. C. (2004). Membrane protein expression and production: effects of polyhistidine tag length and position. *Protein Expr Purif* **33**, 311-25.

Moore, A., Mercer, J., Dutina, G., Donahue, C. J., Bauer, K. D., Mather, J. P., Etcheverry, T. & Ryll, T. (1997). Effects of temperature shift on cell cycle, apoptosis and nucleotide pools in CHO cell batch cultues. *Cytotechnology* **23**, 47-54.

Nakamura, T., Nawa, K. & Ichihara, A. (1984). Partial purification and characterization of hepatocyte growth factor from serum of hepatectomized rats. *Biochem Biophys Res Commun* **122**, 1450-9.

Nakamura, T., Nishizawa, T., Hagiya, M., Seki, T., Shimonishi, M., Sugimura, A., Tashiro, K. & Shimizu, S. (1989). Molecular cloning and expression of human hepatocyte growth factor. *Nature* **342**, 440-3.

Naldini, L., Tamagnone, L., Vigna, E., Sachs, M., Hartmann, G., Birchmeier, W., Daikuhara, Y., Tsubouchi, H., Blasi, F. & Comoglio, P. M. (1992). Extracellular proteolytic cleavage by urokinase is required for activation of hepatocyte growth factor/scatter factor. *EMBO J* **11**, 4825-33.

Nehlsen, K., Schucht, R., da Gama-Norton, L., Kromer, W., Baer, A., Cayli, A., Hauser, H. & Wirth, D. (2009). Recombinant protein expression by targeting pre-selected chromosomal loci. *BMC Biotechnol* **9**, 100.

Nettleship, J. E., Aplin, R., Aricescu, A. R., Evans, E. J., Davis, S. J., Crispin, M. & Owens, R. J. (2007). Analysis of variable N-glycosylation site occupancy in glycoproteins by liquid chromatography electrospray ionization mass spectrometry. *Anal Biochem* **361**, 149-51.

Nettleship, J. E., Assenberg, R., Diprose, J. M., Rahman-Huq, N. & Owens, R. J. (2010). Recent advances in the production of proteins in insect and mammalian cells for structural biology. *J Struct Biol* **172**, 55-65.

Ng, J. D., Lorber, B., Witz, J., Theobald-Dietrich, A., Kern, D. & Giege, R. (1996). The crystallization of biological macromolecules from precipitates: Evidence for Ostwald ripening. *J. Cryst. Growth* **168**, 50-62.

Niemann, H. H., Jager, V., Butler, P. J., van den Heuvel, J., Schmidt, S., Ferraris, D., Gherardi, E. & Heinz, D. W. (2007). Structure of the human receptor tyrosine kinase met in complex with the Listeria invasion protein InlB. *Cell* **130**, 235-46.

O'Gorman, S., Fox, D. T. & Wahl, G. M. (1991). Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251**, 1351-5.

Oumard, A., Qiao, J., Jostock, T., Li, J. & Bode, J. (2006). Recommended Method for Chromosome Exploitation: RMCE-based Cassette-exchange Systems in Animal Cell Biotechnology. *Cytotechnology* **50**, 93-108.

Ozaki, K., Nagata, M., Suzuki, M., Fujiwara, T., Ueda, K., Miyoshi, Y., Takahashi, E. & Nakamura, Y. (1998). Isolation and characterization of a novel human lung-specific gene homologous to lysosomal membrane glycoproteins 1 and 2: significantly increased expression in cancers of various tissues. *Cancer Res* **58**, 3499-503.

Palade, G. (1975). Intracellular aspects of the process of protein synthesis. *Science* **189**, 867.

Palermo, D. P., DeGraaf, M. E., Marotti, K. R., Rehberg, E. & Post, L. E. (1991). Production of analytical quantities of recombinant proteins in Chinese hamster ovary cells using sodium butyrate to elevate gene expression. *J Biotechnol* **19**, 35-47.

Patterson, A. L. (1935). A direct method for the determination of the components of interatomic distances in crystals. *Z. Kristallogr.* **90**, 517-542.

Peng, R.-W. & Fussenegger, M. (2009). Engineering the secretory pathway in mammalian cells. In *Cell line development* (Al-Rubeai, M., ed.), pp. 233-248. Springer Science+Business Media B.V., Dublin.

Petrescu, A. J., Milac, A. L., Petrescu, S. M., Dwek, R. A. & Wormald, M. R. (2004). Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **14**, 103-14.

Pham, P. L., Perret, S., Doan, H. C., Cass, B., St-Laurent, G., Kamen, A. & Durocher, Y. (2003). Large-scale transient transfection of serum-free suspension-growing HEK293 EBNA1 cells: peptone additives improve cell growth and transfection efficiency. *Biotechnol Bioeng* **84**, 332-42.

Pichler, J., Hesse, F., Wieser, M., Kunert, R., Galosy, S. S., Mott, J. E. & Borth, N. (2009). A study on the temperature dependency and time course of the cold capture antibody secretion assay. *J Biotechnol* **141**, 80-3.

Pikaart, M. J., Recillas-Targa, F. & Felsenfeld, G. (1998). Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev* **12**, 2852-62.

Pilbrough, W., Munro, T. P. & Gray, P. (2009). Intraclonal protein expression heterogeneity in recombinant CHO cells. *PLoS One* **4**, e8432.

Porter, K., Prescott, D. & Frye, J. (1973). Changes in surface morphology of Chinese hamster ovary cells during the cell cycle. *J Cell Biol* **57**, 815-36.

Porteus, M. H. & Baltimore, D. (2003). Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763.

Pruett-Miller, S. M., Connelly, J. P., Maeder, M. L., Joung, J. K. & Porteus, M. H. (2008). Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. *Mol Ther* **16**, 707-17.

Puttini, S., Ouvrard-Pascaud, A., Palais, G., Beggah, A. T., Gascard, P., Cohen-Tannoudji, M., Babinet, C., Blot-Chabaud, M. & Jaisser, F. (2005). Development of a targeted transgenesis strategy in highly differentiated cells: a powerful tool for functional genomic analysis. *J Biotechnol* **116**, 145-51.

Qiao, J. (2009). Systematic Study on Generation of Mammalian Production Cell Lines by Targeted Integration (RMCE) Ph.D. thesis, University of Braunschweig

Qiao, J., Oumard, A., Wegloehner, W. & Bode, J. (2009). Novel tag-and-exchange (RMCE) strategies generate master cell clones with predictable and stable transgene expression properties. *J Mol Biol* **390**, 579-94.

Raymond, C. S. & Soriano, P. (2007). High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS ONE* **2**, e162.

Reeves, P. J., Callewaert, N., Contreras, R. & Khorana, H. G. (2002). Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc Natl Acad Sci U S A* **99**, 13419-24.

Rice, P., Longden, I. & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276-7.

Ringrose, L., Chabanis, S., Angrand, P. O., Woodroofe, C. & Stewart, A. F. (1999). Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. *EMBO J* **18**, 6630-41.

Romier, C., Ben Jelloul, M., Albeck, S., Buchwald, G., Busso, D., Celie, P. H., Christodoulou, E., De Marco, V., van Gerwen, S., Knipscheer, P., Lebbink, J. H., Notenboom, V., Poterszman, A., Rochel, N., Cohen, S. X., Unger, T., Sussman, J. L., Moras, D., Sixma, T. K. & Perrakis, A. (2006). Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. *Acta Crystallogr D Biol Crystallogr* **62**, 1232-42.

Sambrook, J. & Russell, D. W. (2000). *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory Press.

Sandig, V., Rose, T., Winkler, K. & Brecht, R. (2005). Mammalian cells. In *Production of recombinant proteins. Novel microbial and eukaryotic expression systems.* (Gallisen, G., ed.), pp. 233-252. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

Sauerwald, T. M., Betenbaugh, M. J. & Oyler, G. A. (2002). Inhibiting apoptosis in mammalian cell culture using the caspase inhibitor XIAP and deletion mutants. *Biotechnol Bioeng* **77**, 704-16.

Schebelle, L., Wolf, C., Stribl, C., Javaheri, T., Schnutgen, F., Ettinger, A., Ivics, Z., Hansen, J., Ruiz, P., von Melchner, H., Wurst, W. & Floss, T. Efficient conditional and promoter-specific in vivo expression of cDNAs of choice by taking advantage of recombinase-mediated cassette exchange using FlEx gene traps. *Nucleic Acids Res* **38**, e106.

Schlake, T. & Bode, J. (1994). Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry* **33**, 12746-51.

Schmidt, E. E., Taylor, D. S., Prigge, J. R., Barnett, S. & Capecchi, M. R. (2000). Illegitimate Cre-dependent chromosome rearrangements in transgenic mouse spermatids. *Proc Natl Acad Sci U S A* **97**, 13702-7.

Schönflies, A. M. (1886). Über Gruppen von Bewegungen. (Erste Abhandlung). *Math Ann* **28**, 319-342.

Schönflies, A. M. (1887). Über Gruppen von Bewegungen. (Zweite Abhandlung). *Math Ann* **29**, 50-80.

Schubeler, D., Maass, K. & Bode, J. (1998). Retargeting of retroviral integration sites for the predictable expression of transgenes and the analysis of cis-acting sequences. *Biochemistry* **37**, 11907-14.

Schucht, R. (2006). Entwicklung von flexiblen Zelllinien für die Produktion rekombinanter Proteine und Retroviren. Ph. D. thesis, University of Braunschweig, Germany

Schucht, R., Coroadinha, A. S., Zanta-Boussif, M. A., Verhoeyen, E., Carrondo, M. J., Hauser, H. & Wirth, D. (2006). A new generation of retroviral producer cells: predictable and stable virus production by Flp-mediated site-specific integration of retroviral vectors. *Mol Ther* **14**, 285-92.

Scrable, H. & Stambrook, P. J. (1997). Activation of the lac repressor in the transgenic mouse. *Genetics* **147**, 297-304.

Seibler, J. & Bode, J. (1997). Double-reciprocal crossover mediated by FLP-recombinase: a concept and an assay. *Biochemistry* **36**, 1740-7.

Seibler, J., Schubeler, D., Fiering, S., Groudine, M. & Bode, J. (1998). DNA cassette exchange in ES cells mediated by Flp recombinase: an efficient strategy for repeated modification of tagged loci by marker-free constructs. *Biochemistry* **37**, 6229-34.

Senecoff, J. F. & Cox, M. M. (1986). Directionality in FLP protein-promoted site-specific recombination is mediated by DNA-DNA pairing. *J Biol Chem* **261**, 7380-6.

Silver, D. P. & Livingston, D. M. (2001). Self-excising retroviral vectors encoding the Cre recombinase overcome Cre-mediated cellular toxicity. *Mol Cell* **8**, 233-43.

Smith, J. R., Maguire, S., Davis, L. A., Alexander, M., Yang, F., Chandran, S., ffrench-Constant, C. & Pedersen, R. A. (2008). Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. *Stem Cells* **26**, 496-504.

Southern, P. J. & Berg, P. (1982). Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J Mol Appl Genet* **1**, 327-41.

Stamos, J., Lazarus, R. A., Yao, X., Kirchhofer, D. & Wiesmann, C. (2004). Crystal structure of the HGF beta-chain in complex with the Sema domain of the Met receptor. *EMBO J* **23**, 2325-35.

Stanley, P. (1989). Chinese hamster ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity. *Mol Cell Biol* **9**, 377-83.

Sternberg, N. & Cohen, G. (1989). Genetic analysis of the lytic replicon of bacteriophage P1. II. Organization of replicon elements. *J Mol Biol* **207**, 111-33.

Stoker, M. & Perryman, M. (1985). An epithelial scatter factor released by embryo fibroblasts. *J Cell Sci* **77**, 209-23.

Strack, R. L., Strongin, D. E., Bhattacharyya, D., Tao, W., Berman, A., Broxmeyer, H. E., Keenan, R. J. & Glick, B. S. (2008). A noncytotoxic DsRed variant for whole-cell labeling. *Nat Methods* **5**, 955-7.

Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D. & Eisenberg, D. (2006). Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **103**, 8060-5.

Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H. & Hung, L. W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* **65**, 582-601.

Thorpe, H. M. & Smith, M. C. (1998). In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci U S A* **95**, 5505-10.

Thyagarajan, B. & Calos, M. P. (2005). Site-specific integration for high-level protein production in mammalian cells. *Methods Mol Biol* **308**, 99-106.

Thyagarajan, B., Guimaraes, M. J., Groth, A. C. & Calos, M. P. (2000). Mammalian genomes contain active recombinase recognition sites. *Gene* **244**, 47-54.

Tjio, J. H. & Puck, T. T. (1958). Genetics of somatic mammalian cells. II. Chromosomal constitution of cells in tissue culture. *J Exp Med* **108**, 259-68.

Tolbert, W. D., Daugherty-Holtrop, J., Gherardi, E., Vande Woude, G. & Xu, H. E. (2010). Structural basis for agonism and antagonism of hepatocyte growth factor. *Proc Natl Acad Sci U S A* **107**, 13264-9.

Toledo, F., Liu, C. W., Lee, C. J. & Wahl, G. M. (2006). RMCE-ASAP: a gene targeting method for ES and somatic cells to accelerate phenotype analyses. *Nucleic Acids Res* **34**, e92.

Tomicic, M. T., Thust, R., Sobol, R. W. & Kaina, B. (2001). DNA polymerase beta mediates protection of mammalian cells against ganciclovir-induced cytotoxicity and DNA breakage. *Cancer Res* **61**, 7399-403.

Turan, S., Galla, M., Ernst, E., Qiao, J., Voelkel, C., Schiedlmeier, B., Zehe, C. & Bode, J. (2011). Recombinase-mediated cassette exchange (RMCE): traditional concepts and current challenges. *J Mol Biol*, doi: 10.1016/j.jmb.2011.01.004.

Turan, S., Kuehle, J., Schambach, A., Baum, C. & Bode, J. (2010). Multiplexing RMCE: Versatile Extensions of the Flp-Recombinase-Mediated Cassette-Exchange Technology. *J Mol Biol* **402**, 52-69.

Urlaub, G. & Chasin, L. A. (1980). Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. *Proc Natl Acad Sci U S A* **77**, 4216-20.

Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D. & Holmes, M. C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646-51.

Vagin, A. A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J Appl Cryst* **30**, 157-163.

Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr D Biol Crystallogr* **55**, 191-205.

Van Craenenbroeck, K., Vanhoenacker, P. & Haegeman, G. (2000). Episomal vectors for gene expression in mammalian cells. *Eur J Biochem* **267**, 5665-78.

Verhoeyen, E. (2000). Efficient targeting of retroviral FRT tagged chromosomal loci and the production of pigs transgenic for human complement regulatory proteins as donors for xenotransplantation. Ph. D. thesis, University of Braunschweig, Germany.

Verhoeyen, E., Hauser, H. & Wirth, D. (2001). Evaluation of retroviral vector design in defined chromosomal loci by Flp-mediated cassette replacement. *Hum Gene Ther* **12**, 933-44.

Waite, L. L. & Cox, M. M. (1995). A protein dissociation step limits turnover in FLP recombinase-mediated site-specific recombination. *J Biol Chem* **270**, 23409-14.

Wallace, H. A., Marques-Kranc, F., Richardson, M., Luna-Crespo, F., Sharpe, J. A., Hughes, J., Wood, W. G., Higgs, D. R. & Smith, A. J. (2007). Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. *Cell* **128**, 197-209.

Walters, M. C., Magis, W., Fiering, S., Eidemiller, J., Scalzo, D., Groudine, M. & Martin, D. I. (1996). Transcriptional enhancers act in cis to suppress position-effect variegation. *Genes Dev* **10**, 185-95.

Weidner, K. M., Arakaki, N., Hartmann, G., Vandekerckhove, J., Weingart, S., Rieder, H., Fonatsch, C., Tsubouchi, H., Hishida, T., Daikuhara, Y. & et al. (1991). Evidence for the identity of human scatter factor and human hepatocyte growth factor. *Proc Natl Acad Sci U S A* **88**, 7001-5.

White, J. F., Trinh, L. B., Shiloach, J. & Grisshammer, R. (2004). Automated large-scale purification of a G protein-coupled receptor for neurotensin. *FEBS Lett* **564**, 289-93.

Whitelaw, E., Sutherland, H., Kearns, M., Morgan, H., Weaving, L. & Garrick, D. (2001). Epigenetic effects on transgene expression. *Methods Mol Biol* **158**, 351-68.

Wilke, S., Krausze, J., Gossen, M., Groebe, L., Jäger, V., Gherardi, E., van den Heuvel, J. & Büssow, K. (2010). Glycoprotein production for structure analysis with stable, glycosylation mutant CHO cell lines established by fluorescence-activated cell sorting. *Protein Sci* **19**, 1264-71.

Wong, E. T., Kolman, J. L., Li, Y. C., Mesner, L. D., Hillen, W., Berens, C. & Wahl, G. M. (2005). Reproducible doxycycline-inducible transgene expression at specific loci generated by Cre-recombinase mediated cassette exchange. *Nucleic Acids Res* **33**, e147.

Wurm, F. M. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol* **22**, 1393-8.

Wurm, F. M., Gwinn, K. A. & Kingston, R. E. (1986). Inducible overproduction of the mouse c-myc protein in mammalian cells. *Proc Natl Acad Sci U S A* **83**, 5414-8.

Yoshikawa, T., Nakanishi, F., Ogura, Y., Oi, D., Omasa, T., Katakura, Y., Kishimoto, M. & Suga, K. (2000). Amplified gene location in chromosomal DNA affected recombinant protein production and stability of amplified genes. *Biotechnol Prog* **16**, 710-5.

Zhang, C. & Kim, S. H. (2003). Overview of structural genomics: from structure to function. *Curr Opin Chem Biol* **7**, 28-32.

Danksagung

Die vorliegende Arbeit wurde unter der Anleitung von Herrn Dr. Konrad Büssow am Helmholtz-Zentrum für Infektionsforschung, Braunschweig angefertigt. Daher gilt mein besonderer Dank Konrad für die Ermöglichung dieser Dissertation und für die ausgezeichnete Betreuung, Unterstützung und stete Bereitschaft zur Diskussion.

Herrn Prof. Dr. Dirk W. Heinz und Herrn Prof. Dr. Stefan Dübel danke ich dafür, dass sie diese Arbeit vor der Naturwissenschaftlichen Fakultät der Technischen Universität Braunschweig vertreten. Herrn Prof. Dr. Michael Steinert danke ich für seine Bereitschaft zur Teilnahme an der Prüfungskommission.

Ein großes Dankeschön geht an Nick Quade, Felix Deluweit, Steffen Meyer, Agnes Zimmer, Lilia Polle, Zhaopeng Li, Uwe Wengler und alle anderen Mitdoktoranden sowie alle Mitarbeiter und Mitarbeiterinnen des Bereichs Molekulare Strukturbiologie für das nette Arbeitsklima und die gute Zusammenarbeit.

Daniela Oster möchte ich von Herzen danken, dass sie mich am Anfang meiner Arbeit eingearbeitet und technisch unterstützt hat. Meiner Diplomandin Agathe Duda und Vitali Maffenbeier sowie den anderen Praktikanten, die mich tatkräftig bei meinem Projekt unterstützt haben, möchte ich ebenfalls danken.

Für ihre Unterstützung und das stete Interesse an meiner Arbeit möchte ich Dr. Joop van den Heuvel und Dr. Volker Jäger ein großes Dankeschön zukommen lassen. Besonders für die Arbeit in der Welt der Zellkulturen und Bioreaktoren.

Für die technische Unterstützung geht mein Dank an Sarah Tokarski und Nadine Konisch.

Besonders möchte ich mich bei Dr. Jörn Krauße, Dr. Joachim Reichelt und Dr. Björn Klink für die praktische Einführung in die Kristallographie bedanken.

Ein wertvoller Beitrag zu dieser Arbeit waren die Gespräche mit Prof. Dr. Jürgen Bode, Dr. Roland Schucht und Dr. Dagmar Wirth. Vielen Dank dafür. In diesem Zuge geht ein herzlicher Dank an Dr. Andre Oumard, Sören Turan, Junhua Qiao und Sandra Broll aus der Arbeitsgruppe von Prof. Dr. Jürgen Bode für die Einführung in die Kunst der Southern Blots.

An Dr. Lothar Gröbe und Maria Höxter geht mein Dank für die zahlreichen Zellsortierungseinsätze.

Mein herzlichster Dank geht an Michael Reck für das Korrekturlesen und vor allem für die wundervolle Zeit zwischen der Arbeit. Ich bin froh, dass es Dich gibt.

Der größte Dank gilt meiner Familie, die mir diese lange Ausbildung ermöglicht hat. Sie hat mich stets unterstützt und somit einen wesentlichen Anteil am Erfolg dieser Arbeit.